Ana Maria D'Ávila Lopes (Brazil)
Sebastian Smart (England)
Keigo Komamura (Japan)
(Editors)

Neurolaw

Legal Impacts of Neurotechnology

Aline Passos Maia
Ana Maria D'Ávila Lopes
Antonio Jorge Pereira Júnior
Camila Pintarelli
Cecília Santana Figueiredo Pinto Alberto
Denise Almeida de Andrade
Esteban Oyarzún Gómez
Felipe Paredes
Francisco Luciano Lima Rodrigues
Janna da Nóbrega Souza
José Luis González Ávila
José Octávio de Castro Melo
Keigo Komamura
Larissa Maciel do Amaral
María Concepción Rayón Ballesteros
Maria Pia Chirinos
Monica Sapucaia Machado
Monique Pyrrho
Murilo Karasinski
Natalia Leonor Monti
Patrícia Moura Monteiro Cruz
Pedro Araújo
Raquel Passos Maia
Renato Cesar Cardoso
Renato Vilardo de Mello Cruz
Sebastián Smart
Tamami Fukushi
Tiago Gagliano Pinto Alberto
Tsung-Ling Lee
Viviane Ceolin Dallasta Del Grossi

livraria DO ADVOGADO editora

LIVRARIA DO ADVOGADO EDITORA
Rua Riachuelo, 1334 s/104
90010-273 - Porto Alegre - RS
Fone: (51) 3225-3311
livraria@doadvogado.com.br
www.doadvogado.com.br

ISBN 978-85-9590-136-0

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico
UNIFOR PPGD Programa de Pós-Graduação em Direito Constitucional

Neurolaw

Legal Impacts of Neurotechnology

International Cataloguing-in-Publication Data (CIP)

N476 Neurolaw : legal impacts of neurotechnology / Ana Maria D'Ávila Lopes, Sebastian Smart , Keigo Komamura (editors) ; Aline Passos Maia ... [et al.]. – Porto Alegre : Livraria do Advogado, 2025.

297 p. ; 25 cm.

Includes bibliography.

ISBN 978-85-9590-136-0

1. Law - Neurolaw. 2. Neuroscience. 3. Neurotechnology - Human Rights 4. Neuroscience - Democracy. 5. Neuroscience - Labor relations. I. Lopes, Ana Maria D'Ávila. II. Smart, Sebastian. III. Komamura, Keigo. IV. Maia, Aline Passos.

CDU 34:616.8

Index to systematic catalog:

1. Law - Neurolaw 34:616.8

(Librarian in charge: Sabrina Leal Araujo – CRB 8/10213)

Ana Maria D'Ávila Lopes (Brazil)
Sebastian Smart (England)
Keigo Komamura (Japan)
(Editors)

Neurolaw
Legal Impacts of Neurotechnology

Aline Passos Maia
Ana Maria D'Ávila Lopes
Antonio Jorge Pereira Júnior
Camila Pintarelli
Cecília Santana Figueiredo Pinto Alberto
Denise Almeida de Andrade
Esteban Oyarzún Gómez
Felipe Paredes
Francisco Luciano Lima Rodrigues
Janna da Nóbrega Souza
José Luis González Ávila
José Octávio de Castro Melo
Keigo Komamura
Larissa Maciel do Amaral
María Concepción Rayón Ballesteros
Maria Pia Chirinos
Monica Sapucaia Machado
Monique Pyrrho
Murilo Karasinski
Natalia Leonor Monti
Patrícia Moura Monteiro Cruz
Pedro Araújo
Raquel Passos Maia
Renato Cesar Cardoso
Renato Vilardo de Mello Cruz
Sebastián Smart
Tamami Fukushi
Tiago Gagliano Pinto Alberto
Tsung-Ling Lee
Viviane Ceolin Dallasta Del Grossi

livraria DO ADVOGADO *editora*

Porto Alegre, 2025

Cover, graphic design and layout of
Livraria do Advogado Editora

Cover image of
https://www.canva.com/design/

Financing
UNIFOR – Edital do Programa de Apoio às Equipes de Pesquisa nº 61/2023
Chamada CNPq no 14/2023

Edition of
Comércio de Livros dos Advogados Ltda.
Livraria do Advogado Editora
Rua Riachuelo, 1334 s/104
90010-273 Porto Alegre RS
Fone: (51) 3225.3311
livraria@doadvogado.com.br
www.doadvogado.com.br

Printed in Brazil

— Preface —

The headlong advance of Neurotechnology brings countless benefits to humanity across a wide range of areas. However, alongside with these benefits, the potential of Neurotechnology to violate human rights has also been growing. Harmonizing techno-scientific development with respect for human dignity is a crucial challenge nowadays, given that we are facing advances whose impacts may not only increase inequality among human beings but may also be irreversible.

This challenge was the subject of academic debates at the III Seminar of the International Network on Neurolaw and Human Rights (www.neurorights.com.br), held at the University of Fortaleza (Brazil) on October 28, 29 and 30, 2024. The Network, officially launched in October 2022, is composed of researchers from the University of Fortaleza (UNIFOR), Austral University of Chile (UACH), The Externado University of Colombia (UExternado) and Pontifical Catholic University of Rio Grande do Sul (PUCRS). The Network's activities are currently funded by UNIFOR, through the Call for Proposals for the Support Program for Research Teams n. 61/2023 of the Vice-Rectory of Research of the University of Fortaleza (VRP/UNIFOR), as well as by the National Council for Scientific and Technological Development (CNPq), a Brazilian federal agency for research promotion, through the Call CNPq/MCTI/FNDCT n. 18/2021 and the Call for Support for International Scientific, Technological and Innovation Research Projects n. 14/2022.

The III Seminar, held thanks to the financial support of Call n. 37/2023 (PAEP) of the Coordination for the Improvement of Higher Education Personnel (CAPES), another Brazilian federal agency for research promotion, is one of the Network's academic activities. The Network also conducts research and teaching activities in the field of Neurolaw, in partnership with other Brazilian and international academic institutions.

At the 2024 Seminar, the chosen theme was "Neurotechnology: the Future of Humanity", with the participation, both in person and virtually, of more than 30 researchers from different countries, notably highlighting the participation of Professor Nita Farahany, from Duke University (United States), who gave the keynote conference.

This book brings together some of the lectures given during the III Seminar as a contribution to the academic debate and to support legislative initiatives and judicial decisions that have been growing significantly. To make it easier to read, the book is divided into five thematic parts.

The first part, entitled "Neuroethics: Perspectives of Neurotechnology Advances", brings together five chapters. The first is by Professor María Pía Chirinos from

the University of Piura (UDEP) in Peru, entitled "The bodily dimension as the Achilles heel of neuroethics". In this chapter, the author exposes the tenuous boundary between the corporeal and the rational, an essential theme for a neuroethical perspective on advances in Neurotechnology. The second chapter, entitled "The belief in a just world: neuroscientific perspectives and legal implications", is by Professors Murilo Karasinski, from the Pontifical Catholic University of Paraná (PUCPR) in Brazil, and Renato Cesar Cardoso, from the Federal University of Minas Gerais (UFMG) in Brazil, in which they analyze how neuroscience has explored the structuring of ways to interpret everyday events and the judge behaviors that shape perceptions of merit and responsibility, based on the belief in a just world. The third chapter is by Professor José Octávio de Castro Melo from the State University of Piauí (UESPI) in Brazil, and is entitled: "Paradoxes of Transhumanism: Impacts on Equality, Freedom, and Human Dignity", in which he seeks to answer to what extent transhumanism can impact the principles of equality, freedom, and human dignity, considering the challenges of accessibility and the risks of worsening social inequalities in a context of rapid technological advancement. The fourth chapter is by Professor Monique Pyrrho and master's candidate Pedro Araújo from the University of Brasília (UNB) in Brazil, entitled "Neurotechnology, ethics and human rights: Thinking about work, mental health inequities and eugenics", in which they aim to analyze the ethical challenges of neurotechnologies considering their risks, especially regarding relation to increasing inequalities among human beings. In the fifth and final chapter of this first part of the book, Professor Tamami Fukushi from Tokyo Online University (TOU) in Japan, in her chapter entitled "The Future of Neurotechnology: Co-evolution between Neuroethics and Neurotechnology", analyzes the history of neuroethics, describes the development of Neurotechnology and discusses the future ethical concerns that it raises.

In the Second Part of the book, "Politics and Economic Impacts of Neurotechnology", there are two chapters. The first, entitled "Neurolaw in Non-Democratic Countries", is by Professor Tiago Gagliano Pinto Alberto and student Cecília Santana Figueiredo Pinto Alberto from the Pontifical Catholic University of Paraná (PUCPR) in Brazil. Theyseek to analyze the delicate balance between technological evolution and freedom of thought, aiming specifically to elucidate to what extent privacy and mental integrity are guaranteed both internally and externally in non-democratic regimes. Their analysis serves as a focal point for understanding the distinction between these regimes and democratic states, their peculiarities and their relationship with freedom of thought. In the second chapter, "Abuse of economic power in the age of neurotechnology", the State Attorney of São Paulo (PGE/SP) of Brazil, Camila Pintarelli, aims to demonstrate the emergence of a new form of informational asymmetry derived from access to neural data, highlighting the need to broaden the legal understanding of the abuse of economic power, so that it can also encompass, even hermeneutically, the violation of cognitive freedom and mental privacy.

The third part of the book, which revolves around the theme of "Legal Regulation of Neurotechnology", comprises five chapters. In the first, "Governing Neurotechnology Globally: Between Norms Fragmentation and Coherence", Professor Tsung-Ling Lee at Taipei Medical University (TMU) in Taiwan, examines how the fragmentation of norms on Neurotechnology occurs as several international organizations with partially overlapping mandates introduce guidance documents and prin-

ciples motivated by varied concerns and interests. In the second chapter, entitled "Emerging technologies and protection of human rights: applying the UNGPs in the neurotechnology field", the Federal Public Defender (DPU) of Brazil, Viviane Ceolin Dallasta Del Grossi, aims to identify due diligence mechanisms used by technology companies to mitigate risks and prioritize human rights in practice, noting the regulatory legislative vacuum that still prevails on the subject. The third chapter, "The hardening of international soft law documents in the field of Neurolaw", is by Professor Ana Maria D'Ávila Lopes of the University of Fortaleza (UNIFOR) in Brazil, who proposes the strengthening of existing international soft law standards, as a way to overcome the vacuum in national legal systems on Neurotechnology. The fourth chapter is by Professor Sebastián Smart from Anglia Ruskin University (ARU) in England, and lawyer Esteban Oyarzún Gómez, director of International Human Rights Study Circle (CEDIDH) of Chile. In this chapter they analyze strategic litigation in human rights (across different fields), adapted to the field of Neurotechnology and artificial intelligence. In the fifth and final chapter of this part of the book, lawyer Natalia Leonor Monti from Foundation Kamanau, based in Argentina, aims to demonstrate the need to protect democratic systems in the face of new challenges imposed by scientific advances in Neurotechnology and its convergence with Artificial Intelligence and immersive technologies. She focuses on the Americas, highlighting the region's leadership in regulatory initiatives on the subject.

The fourth part of the book, "Neurotechnology and Minorities (De)protection", comprises four chapters. In the first, entitled "Neurotechnological advancements for fairness: detecting and mitigating implicit racial bias in criminal sentencing", Professor Janna da Nóbrega Souza, from the University of Murcia (UM) in Spain, aims to demonstrate the need to rigorously analyze the impact of neuroscientific advances on the judicial process, in order to avoid racially discriminatory decisions. In the second chapter, "Gender-based algorithmic discrimination and just transition in the framework of the fourth industrial revolution and technological innovation", by professors Denise Almeida de Andrade, from the Christus University Center (UNICHRISTUS) in Brazil, and Monica Sapucaia Machado, from the Institute of Education, Development and Research (IDP) in Brazil, argue for the need critically examine algorithmic discrimination, which is a present reality and disproportionately affects vulnerable groups, exacerbating inequalities, particularly gender discrimination. The third chapter, "Child Influencers and the Neuro Rights of Children and Adolescents: An Analysis of Vulnerability in the Digital Environment" was written by Professor Patricia Moura Monteiro Cruz of the University of Fortaleza (UNIFOR) in Brazil, the State Attorney of Ceara (PGE/SP) of Brazil, Renato Vilardo de Mello Cruz, and Professor Antonio Jorge Pereira Júnior from the University of Fortaleza (UNIFOR) in Brazil. This chapter aims to highlight the implications of children and adolescents' roles as digital influencers from the perspective of neurorights, investigating how these can be applied to provide more effective protection in the digital environment. The fourth chapter, "Veripol: legal and technical analysis of the tool for the detection of false reports" by professors María Concepción Rayón Ballesteros from the Complutense University of Madrid (UCM) in Spain, and José Luis González Ávila from the University of La Laguna (ULL) in Spain, conducts a technical and legal study of VERIPOL, an artificial intelligence tool used by security forces in Spain to detect false reports of crimes, mainly related to robberies and thefts.

The Fifth Part of the book "Updating Neurorights" has three chapters. The first, entitled, "The automation of the Judiciary: to what extent new technologies can put fundamental and social rights at risk. The centrality of the person" is written by Professor Francisco Luciano Lima Rodrigues from the University of Fortaleza (UNIFOR) in Brazil. He proposes a critical analysis of the automation of the Judiciary, emphasizing the need for the use of technology to respect the centrality of the person and not be solely focused on achieving institutional efficiency. In the second chapter, "Neural data protection and the need to update the self-management model of privacy", Professor Felipe Paredes from the Austral University of Chile (UACH) in Chile, argues for the need to update current privacy protection mechanisms in order to adequately protect the cognitive processes of individuals, while avoiding the creation of a new right, specifically to protect brain privacy. In the third and final chapter of this part, "The Commodification of Free Will: Neurotechnology and Ethical Challenges in the Workplace," Professor Aline Passos Maia from the University of Fortaleza (UNIFOR) in Brazil, PhD candidate Raquel Passos Maia from the University of Lisbon (ULISBOA) in Portugal, and Professor Larissa Maciel do Amaral from the University of Fortaleza (UNIFOR) in Brazil discuss how the commodification of neural data introduces profound ethical dilemmas in the workplace, such as whether employers can claim ownership over the cognitive processes of their employees', highlighting the emergence of a new standard for fundamental human rights..

Finally, Professor Keigo Komamura from Keio University (KEIO) in Japan concludes the book with a brief perspective on the movement that gave rise to a new legal branch called "Neurolaw."

In this way, the book hopes to contribute to the necessary academic dialogue on the ethical and legal limits of advances in Neurotechnology, aiming to safeguard the future of humanity.

Ana Maria D'Ávila Lopes

Professor at the University of Fortaleza (Brazil)
CNPq Research Productivity Fellow
Coordinator of the International Network of Neurolaw and Human Rights

— About the authors —

ALINE PASSOS MAIA

Professor of Law at the Center of Legal Sciences of the University of Fortaleza (Brazil). PhD in Labor Law from the University of Salamanca, Spain. Master's degree in Labor Law from the University of Salamanca, Spain. alinepmaia@unifor.br

ANA MARIA D'ÁVILA LOPES

Full Professor of Law at the University of Fortaleza (Brazil). Post-Doctorate in Law from the University of Auckland. PhD in Law from the Federal University of Minas Gerais. Research Productivity Fellow (PQ/CNPq-Brazil). anadavilalopes@unifor.br

ANTONIO JORGE PEREIRA JÚNIOR

Full Professor of Law at the University of Fortaleza (Brazil). PhD. in Law from University of São Paulo. Jabuti Award Winner - Law Category (2012). antoniojorge2000@gmail.com

CAMILA PINTARELLI

State Attorney of São Paulo (Brazil). Advocate for the constitutional reform of neurorights in Brazil. Post-Doctorate in Law from the University of Coimbra. PhD in Law from the Pontifical Catholic University of São Paulo. ckpintarelli@gmail.com

CECÍLIA SANTANA FIGUEIREDO PINTO ALBERTO

Undergraduate student at Pontifical Catholic University of Paraná (Brazil).
ceciliasfigueiredo2507@gmail.com

DENISE ALMEIDA DE ANDRADE

Professor of the Academic Master's and Undergraduate Law Programs at the Christus University Center (Brazil). Post-doctorate in Political and Economic Law from Mackenzie Presbyterian University. PhD and Master's Degree in Constitutional Law from the University of Fortaleza . Professor at the Getúlio Vargas Foundation - FGVLaw São Paulo. Visiting Professor at Shepherd University - West Virginia (USA). andradedenise@hotmail.com

ESTEBAN OYARZÚN GÓMEZ

Bachelor's degree in Legal and Social Sciences from the University Austral of Chile (Chile). Director of the International Human Rights Studies Circle (CEDIDH) of Chile. eoyarzungomez@hotmail.com

FRANCISCO LUCIANO LIMA RODRIGUES

Full Professor of the Master's/Doctorate Graduate Program in Constitutional Law at the University of Fortaleza. Associate Professor at the Faculty of Law of the Federal University of Ceará. Judge of the Court of Justice of the State of Ceará. PhD in Law from the Federal University of Pernambuco, with a research internship at the Faculty of Law of the University of Lisbon Portugal. lucianolima@unifor.br

FELIPE PAREDES

Associate Professor at the Austral University of Chile (Chile). Master and PhD in Law from the Pompeu Fabra University (Spain). felipe.paredes@uach.cl

JANNA DA NÓBREGA SOUZA

Professor of Criminal Law at the University of Murcia (Spain). PhD from the University of Murcia. Master's degree from the Carlos III University of Madrid. janna.nobrega@um.es

JOSÉ LUIS GONZÁLEZ ÁVILA

Professor of the Computer Science Department of the University of La Laguna (Spain). Civil servant of the Autonomous Community of the Government of the Canary Islands, Project Manager of the Platform for Processing of Administrative Procedures. Professor-Tutor of Industrial Engineering at the National University of Distance Studies. jlgavila@ull.edu.es

JOSÉ OCTÁVIO DE CASTRO MELO

Assistant professor at the State University of Piaui (Brazil). PhD. Candidate in Constitutional Law of the University of Fortaleza. Master's degree in theory and philosophy of law from the Federal University of Pernambuco. Specialized in Procedural Law from the Federal University of Santa Catarina. Bachelor's degree in law from the Federal University of Piaui. joseoctavio@edu.unifor.br

KEIGO KOMAMURA

Professor of the Faculty of Law of Keio University (Japan). Master and PhD of Law from Keio University. komamura@keio.jp

LARISSA MACIEL DO AMARAL

Professor at the University of Fortaleza and Christus University Center (Brazil). Master in law from the Federal University of Ceará. PhD in Business Administration from the University of Fortaleza. Lawyer at Rabb Carvalho Advocacia. larissaamaral@unifor.br

MARÍA CONCEPCIÓN RAYÓN BALLESTEROS

Professor of Law at the Complutense University of Madrid (Spain). Lawyer of the Bar Association of Madrid. Mediator registered with the Ministry of Justice. PhD in Law and Licentiate in Political Science and Administration. mcrayon@ucm.es

MARIA PIA CHIRINOS

Professor at the Faculty of Humanities at the University of Piura (Peru). PhD in Philosophy at the University of Navarra and Vice-Rector at the University of Piura - Lima Campus. mariapia.chirinos@udep.edu.pe

MONICA SAPUCAIA MACHADO

Professor and Head of the Master's Program in Law, Justice, and Development at IDP. Member of the National Education Council (CNE) of the Brazilian Federal Government. Ph.D. in Political and Economic Law from Presbyterian University Mackenzie. Lawyer (Brazil) msapucaia@yahoo.com

MONIQUE PYRRHO

Professor at the International Center for Bioethics and Humanities at the University of Brasilia (Brazil). PhD in Bioethics from the University of Brasilia. pyrrho.monique@gmail.com

MURILO KARASINSKI

Professor at the Pontifical Catholic University of Parana (Brazil). Master and PhD in Philosophy, both from Pontifical Catholic University of Parana. k.murilo@pucpr.br

NATALIA LEONOR MONTI

Lawyer (Argentina), currently Legal Director of the Kamanau Foundation. nmonti@gmail.com

PATRÍCIA MOURA MONTEIRO CRUZ

Professor at the University of Fortaleza (Brazil). Master and Doctoral candidate in Constitutional Law at the University of Fortaleza. patriciamoura@unifor.br

PEDRO ARAÚJO

Lawyer for representative entities of employees of the Federal Public Prosecutor's Office (Brazil). Master's degree in Bioethics from the University of Brasília. Postgraduate in Criminal Law and Criminal Procedure from Estacio University Center. pedro.jurid@gmail.com

RAQUEL PASSOS MAIA

PhD candidate in Law at the Faculty of Law of the University of Lisbon (Portugal). Master in Constitutional Law at the Faculty of Law of the University of Lisbon. Scholarship holder of the Foundation for Science and Technology-FCT. raquelpmaia13@gmail.com

RENATO CESAR CARDOSO

Associate Professor at the Faculty of Law of the University Federal de Minas Gerais (Brazil). Post-Doctorate in Philosophy from the University of Barcelona. PhD in Philosophy of Law from University Federal de Minas Gerais. Deputy Coordinator of the Postgraduate Program in Neuroscience at University Federal de Minas Gerais. renatoccardoso@hotmail.com

RENATO VILARDO DE MELLO CRUZ

State Prosecutor of Ceará and Lawyer (Brazil). Master's Candidate in Law and Conflict Management at the University of Fortaleza. renatovmcruz@gmail.com

SEBASTIÁN SMART

Senior Research Fellow in Access to Justice, Law and Technology at Anglia Ruskin University (England). PhD in Latin American Studies and Human Rights from University College London, as well as an MA in Human Rights from the same institution. sebastian.smart@aru.ac.uk

TAMAMI FUKUSHI

Professor at the Faculty of Human Welfare of the Tokyo Online University (Japan). PhD in Behavioral Science from the Hokkaido University Japan. BA in Physical Education from the Nara Women's University Japan. fukushi.tamami@internet.ac.jp

TIAGO GAGLIANO PINTO ALBERTO

Professor at the Pontifical Catholic University of Parana (Brazil). PhD in Law from the Federal University of Paraná . Judge of Law of the Fourth Appeals Court of the Court of Justice of the State of Paraná. tiagogagliano@hotmail.com

TSUNG-LING LEE

Associate Professor of Law at Taipei Medical University (Taiwan). Graduate Institute of Health and Biotechnology Law, Taipei Medical University. Master and Doctor of Juridical Science from Georgetown University Law Center (USA), and a Bachelor of Medical Science from the University of Sydney (Australia). tl265265@tmu.edu.tw

VIVIANE CEOLIN DALLASTA DEL GROSSI

Federal Public Defender (Brazil). PhD in Human Rights from the University of São Paulo. Master in Human Rights from the University of São Paulo. Master in Criminology and Criminal Enforcement from Pompeu Fabra University. viviane.dallasta@dpu.def.br.

— Contents —

— IV —

NEUROTECHNOLOGY AND MINORITIES (DES)PROTECTION

— V —

UPDATING NEURORIGHTS

— EPILOGUE —

— I —

NEUROETHICS PERSPECTIVES OF NEUROTECHNOLOGY ADVANCES

— 1 —

The bodily dimension as the Achilles heel of neuroethics[1]

MARIA PIA CHIRINOS

Universidad de Piura (Peru)
mariapia.chirinos@udep.edu.pe

Contents: 1. The status of the body in modern philosophy; 2. Is it still persuasive the Modern vision of matter?; 3. Neuroscience and neuroethics: another perspective on the value of the body; 3. Corporeality and sociability; 4. Final thoughts; References.

1. The status of the body in modern philosophy

The word 'dignity' is clearly a modern concept. It first appeared in the treatise *De hominis dignitate,* by Giovanni Pico della Mirandola (2016) in 1486. This work was written in an explicitly Renaissance context, with a focus on a new anthropocentric ideal of freedom and on the capacity for self-determination that places humans far above all other nature.

Immanuel Kant (1988) espoused the notion of the inherent dignity of the individual, González, 2004) and non-living matter, that is, inorganic matter. Nevertheless it is imperative to comprehend Kant within the context of his historical and philosophical era, particularly in contrast to the moral sense school, which posits that feelings and empirical principles underpin moral law, a doctrine to which Kant had formerly adhered.

Indeed, rationalism is concerned with reason and reason alone, whilst empiricism focuses on sensibility. According to this bifurcation, Kantian reason is termed "pure reason", in the sense that it excludes any sensible or corporeal intervention. An analysis of this condition of "purity" is therefore warranted. The first point to be noted is that the concept of "pure reason" is not a totally modern invention. When exploring the two most renowned definitions of the human being proposed by Aristotle – the 'rational animal' and the 'political animal' – philosophers have predominantly emphasized the most salient aspects of these definitions, namely, rationality

[1] I am grateful to the *De Nicola Center for Ethics and Culture* and the *Jaques Maritain Center*, both at the University of Notre Dame, Indiana, for providing the facilities for this research.

and politicity, disregarding the term 'our animality' (MacIntyre, 1999). The question arises: why has this occurred? The reason for this is that our animality reflects our corporeality, a concept that has been disregarded not only by rationalism, but also by a philosopher who is not modern in his outlook, such as Plato. The body is the prison of the soul and therefore does not appear to provide any essential plus to our humanity; indeed, quite the contrary. Our animal condition would divert us from the purity of our essence and make us resemble a realm plagued by changes. Modernity takes up this Platonic definition of the human being as "soul" or reason, albeit with different nuances. The Aristotelian body as living matter was forgotten or even denied.

2. Is it still persuasive the Modern vision of matter?

According to Descartes, the modern proposal has eliminated a significant portion (if not the entirety) of the organic world in order to expand the inorganic and reduce it to inert matter. This matter has been subjected to the typical necessity of abstract space and time and, consequently, to strict mechanical functions (Descartes, 2010). The only living reality for Descartes was that of the human, in terms of a rational soul accidentally united to the body. The corporeal condition of living beings began to be understood as a passive, external, and mechanical dimension, and thus the corporeal was excluded from the definition of life and began to be overshadowed by the definition of dignity. But is it an innocuous decision for philosophy or for law to deny the status of life to the plant and animal kingdoms, that is, to reduce them to the realm of inorganic material?

In order to gain a more profound understanding of this dilemma, it is helpful to turn to the insights of Fernando Inciarte (1979). Inciarte's work offers a critique of the modern thesis that posits the absence of plant or animal life. However, the most common experience of a living being – whether in the vegetable, animal or rational kingdom – suggests that such a being is born, nourishes, grows, reproduces and dies. This is to say, it is a being with life, as defined by Aristotle in terms of self-movement (1995b, 402a 23–27).

It is important to recognize that, in addition to life's inherent self-movement, many of its corporeal manifestations are truly unique. Metabolism and reproduction are functions of living things that cannot be copied by artifacts or technological models. I am referring more precisely to the ability to recognize inorganic elements from the environment and to assimilate or transform it into one's own corporeal and living substance, or to reproduce a being with the same essential characteristics as ours and open to our own capacities.

This concept of life suggests that the manifestation of life in living beings is not solely attributable to physical organs as such. Indeed, as Robert Spaemann asserts, the satisfaction of bodily needs in living beings signifies transcendence over them, thereby rendering the act of eating profoundly vegetal or animal (that is, according to their specific acts of nutrition and their physical organs). Plants assimilate elements of the earth continuously, and animals eat what they find or kill when they are hungry, without any moral consideration: in an instinctively way. Human beings, on our side, satisfy our corporeal needs freely and non-instinctive. Consequently, Spaemann concludes that humans are neither purely spiritual beings nor disembodied subjec-

tivity capable of disposing of their bodies as if they were mere instruments. In fact, Spaemann asserts that "the human body is man himself (2000, p. 248)." This statement can be referred to plants and animals, too, and it is at odds with the rationalist approach espoused by Descartes, as well as Kant's endeavor to ground dignity exclusively in pure reason.

To illustrate this point related to our human body, we can continue with the example of nutrition, of the seemingly quotidian act of eating. Despite our inherent natural necessity, human beings exceed significant constraints imposed by our physical needs. Not only we transform food into our own organism, but in the act of nutrition, we also introduce characteristics that are exclusive to our rationality. Specifically, we are capable of in-corporate something basic and material like food and adding elements that could be termed superfluous. While theoretically capable of subsisting eating uncooked elements, human beings have opted to alter raw food with fire. In addition, they infuse these materials with flavor, present food in a variety of ways, and share it with others during social gatherings. Human beings humanize food and the act of eating through education and manners, embellishing them with infinite variety. Furthermore, human beings possess the capacity to retain information regarding their dietary intake, to refine cooking techniques and eating habits, to innovate, to create and to perpetuate traditions surrounding food, as well as in artistic and ceremonial contexts. Food has the ability to serve as a medium for conveying emotions, historical significance and even religiosity, as evidenced by numerous ritual practices. The consumption of food and beverages can serve as a medium for reflecting on our rationality, our capacity for social interaction and our sense of dignity.

At this point, however, we encounter other problems. The first concerns the emergence of neuroethics and its postulates about the role of the brain in human behavior. Apparently taking a position contrary to that of modernity, neuroethics seems to affirm the dependence of our actions (and our dignity) on the bodily dimension and, more specifically, on the brain. However, and this would be a first question, are we not faced again the separation of soul and body of the rationalists, only in favor of matter? Can the debate on bodily life help us to answer this question?

3. Neuroscience and neuroethics: another perspective on the value of the body

It is widely acknowledged that the field of neuroethics officially emerged in 2002, following the *Neuroethics: Mapping the Field* event in San Francisco. In that same year, Adina Roskies outlined the distinction so widely accepted today between the ethics of neuroscience, referring to the ethical framework of neuroscientific research, and the neuroscience of ethics, which would deal with the neural basis of moral agency (Roskies, 2002; Cortina, 2010). Adela Cortina has proposed what she terms the *topoi* of neuroethics (Cortina, 2013), which are theses that have gained significant traction among its primary scholars. These *topoi* represent the contribution of neuroscience to the understanding of morality and will be discussed briefly and concisely:

a) A correlation exists between specific brain regions and our moral reasoning, as evidenced primarily through neuroimaging techniques.

b) Due to evolution, we have "moral" codes imprinted in our brains, which since the time of hunter-gatherers, prescribe the defense of the group, of those close to us, etc.

c) Morality is an adaptive mechanism, a set of values, virtues and norms that help us adapt and survive. Moral judgments are instinctive, and their formulation is largely influenced by emotions.

d) Despite the common insistence on the notion that economic rationality, or the maximization of profits, is an innate characteristic of the human species, the field of neuroethics has revealed that all human beings possess a biological predisposition towards cooperation.

e) Human beings are indisputably dependent on their social environment, and to a lesser extent on their genes. The human brain is inherently social due to its capacity for plasticity.

The first three topoi refer to the relationship between morality and the brain (ethics and neurology). And the last two, the relationship between sociability and the brain.

It is conceivable that only a limited number of additional *topoi* could be identified at present; however, these are already adequate to undertake an initial exploration of the role of the body, and more specifically, that of the brain, in ethics. This preliminary affirmation coincides with the relegation of purely rational positions to a secondary status, a phenomenon that has been observed since the 19th century with materialist philosophers like Feuerbach, Marx, etc. and the development of science. Notable events such as the incident involving Phineas P. Gage in 1848 (Kotovicz, 2007) have facilitated substantial, albeit gradual, progress in the field of neuroscience. These advancements have substantiated the correlation between the brain and, most notably, the existence of emotions (which are derived from neuronal activity) as a contributing factor to moral behavior. Indeed, this may be considered one of the central tenets of neuroscience, namely, the identification of the cerebral underpinnings of emotions, thereby providing a definitive explanation of their existence and influence on moral life, as has been asserted from Aristotle (1995a).

The hypothesis that moral codes are embedded within the brain, originating from emotional responses that predate the development of reason, could substantiate the notion that emotions are an integral component of the process of hominization. This process signifies the evolutionary developments inherent to the human body. Consequently, the notion that emotions are intrinsic to the human bodily condition is corroborated, and the assertion that emotions are "the voice of the body" is reinforced. Aristotle would agree, too.

But can we reduce behavior to emotions and emotions to brain's activities? Intriguingly, scholars such as Rizzolatti (2008), the discoverer of mirror neurons, have proposed the innatism of neurological processes as the foundation of ethical judgments. This neuroethical proposal is thus subject to the Cartesian mind-body dichotomy. According to Rizzolatti, innate ideas are formed in the brain, while Descartes' innate ideas pertain to the concept of the soul. Nevertheless, the notion that moral codes are imprinted in the brain, seems not unequivocally substantiated, because the question of the correlation between brain areas and behavior remains unresolved, and the concept of 'correlation' itself is ambiguous. In this case, Aristotle would not agree.

When neuroscience discusses innate concepts, it also introduces a form of determinism. However, as Cortina elucidates, "the evolution of our biological dispositions (...) does not coincide with moral progress at the cultural level" (2013, p. 140). Moreover, the notion of indefinite or irreversible moral or cultural progress is illusory, as recent and not-so-recent historical events demonstrate. At this juncture, Cortina's assertion is noteworthy and aligns with the aforementioned distinction: it is crucial to discern between the cerebral underpinnings of moral behavior, that is, in Inciarte's terminology, the presence of an organ that is at the basis of a function, and the activity in itself — the function — which necessitates the organ but is not reducible to matter. Aristotle's philosophical framework distinguishes between the initial act of a living being, defined as psyche or soul, and subsequent functions that depend on this initial act, including nutrition, reproduction, and knowledge, among others. But at the same time, all those functions do need organs to develop as secondary acts. Organs are necessary conditions, but not sufficient ones.

This prompts further consideration of the human condition, wherein our rational soul and its actions are supported by our body and organs, thus establishing a living unity. As previously discussed, instinctive behavior is associated with univocal responses. However, experience shows that neither our behavior nor our responses are univocal. We can even suspend our moral judgements due to a need for additional information, a preference for alternative approaches, or a desire for more time to contemplate the various aspects of a situation. It is challenging to assert that our judgements are instinctive when their predictability is so problematic. Just as our responses to bodily needs vary significantly, even when these needs are bodily needs, our moral judgements demonstrate comparable variability, despite the emotions that influence them within. Our body and our brain evade the mechanist model due to our freedom. However, this freedom is not absolute, as we are contingent on our bodily condition in numerous ways.

The human condition, in many of its most common yet significant manifestations, exhibits an inherent interconnectedness between our bodily condition and the rational soul, a singularity that is particularly evident in the domain of morality. For Aristotle, morality is the result of the "dialogue" between nature and reason, where "nature" stands for the tendential and orectical dimension of human life. This dimension is not, for Aristotle, absolutely corporeal or without reference to reason (1995a). As González states, these tendencies have to be rationally integrated: it is precisely in such integration that morality consists (2004).

Therefore, a mere comprehension of the neurological underpinnings of emotions and their moral implications from a purely materialistic standpoint is insufficient to elucidate human behavior in its quotidian manifestations, a limitation that is even more pronounced in circumstances where moral judgment pertains to matters of greater import.

3. Corporeality and sociability

Another intriguing topic that emerges among the *topoi* is that of sociability, a concept that has been extensively researched by prominent authors such as Patricia Churchland (2011). It is noteworthy that this relational dimension has already

attained a significant position in the realm of neuroethics, and scientific findings indicate that individualism cannot serve as a framework for interpreting our behavior. Modern thought is once again confronted with a challenge, and it merits further elaboration. The question arises: Are we social animals because we are corporeal? Broadly speaking, the concept of "sociability" refers to the reciprocal relationships between human beings and the subject-subject relationship. The Aristotelian concept of the "political animal" is predicated on the premise of receiving the gift of speech (logos) and the ability to share it with one's equals (Aristotle 1995 c), book 2). However, communication can be not only verbal, but also corporal.

In his seminal work, Alasdair MacIntyre explores the concept of the human body as a medium of expression, asserting that the essence of the human body is defined by its social interconnections with other bodies (2006). Consequently, from the moment of its visibility to others—be it birth or any preceding moment—the human being does not manifest as a political animal that engages in verbal communication, but rather, it does so through bodily expression. And these physical expressions, as expressions for example of newborn babies, can only evoke caring responses from adults, even if they are not their parents.

Neuroethics is a field that encompasses a series of neuroendocrinological and evolutionary biology studies, which provide precise explanations for the presence of reciprocal mutualistic care behaviors among animals. Specifically, Churchland (2011) defends the moral status of animals on the basis of their capacity to care for other animals of their species. In the case of humans, however, we observe not only the capacity to care for our offspring, but also for the offspring of other humans, as well as other species and nonliving beings. Furthermore, humans exhibit the capacity for indirect cooperation (Cortina and Conill, 2019) as evidenced by their concern for universal issues, such as those articulated in the 1947 Declaration of Human Rights and the Sustainable Development Goals, which are similarly oriented towards human rights.

A notable distinction between human and animal caregiving emerges in the following aspect. While certain caring behaviors related to one's own species have been observed among animals, as previously mentioned, a distinguishing feature of human behavior is the presence of a rejection, absence, or abuse of caring acts among the same human beings, which is not observed in animals. Within the human realm, such occurrences are not uncommon.

The rise of environmentalism, with its profound influence on culture, politics, and economics, is often rooted in an exaggerated sense of human capacity to dominate, destroy, or mistreat nature. However, this is the modern meaning of domination and power, which can be found, for example, in Galileo when he writes: "If nature did not answer our questions voluntarily, but hid its secret from us, we would subject it to torture and extract from it, in an agonizing inquisition, the answer that it would not otherwise give" (quoted by Benedict XVI, 1995, p. 24). Technology has gone deeper in this attempt, and the last century has witnessed terrible consequences not only of domination, but also of destruction and human degradation, as notable voices, such as Martin Heidegger (1977) or Hans Jonas (1984), have denounced.

This perspective undergoes a transformation when the right place and relevance of living beings is accurately discerned, encompassing their materiality. This mate-

riality is not merely reduced to pure matter but rather endowed with its own unique function (secondary act) and self-movement. Hence the importance of recognizing that plants and animals, as well as the human body, are living material realities, organic, with their own movement and acts that require organs to live. This transformation is particularly salient in the context of human beings, given our moral and free behavior.

It is important to note that social relationships are not predetermined; individuals have the capacity to establish or disrupt connections. In our case, emotions are imbued with rationality, too, as we have already seen. And this means that, although it is true that emotions can have their origin in the brain and that we, as a free beings, can accept them and direct them towards a specific end, we can also reject them, and even let ourselves be totally carried away by them. In this last case, human being manifest behaviors that are often not typical of the political and rational animal. We are unique animals that can go against our natural inclinations because we can use our freedom even in that sense. And this is also true in the case of our behavior towards other human beings. The social dimension of human beings is a complex and malleable topic, capable of enhancement or deterioration.

4. Final thoughts

The primary challenge confronting neuroscience is to traverse the tenuous boundary between the corporeal and the rational/free. This endeavor can be facilitated by exploring fundamental philosophical concepts such as "life" and its components, including "living matter" and "living brain." In this regard, the German distinction between "Leib" and "Körper" offers a fruitful approach to comprehending the self-movement of the "Leib" that is alive and finalized and not predetermined (as "Körper" is). The advent of the primary act—the act of the soul—in matter, or the living body, has the potential to profoundly impact themes of considerable anthropological significance. In this sense, neuroethics aspires to universal ethics, yet it reduces our moral conduct to its cerebral foundations. This calls into question deeply human issues such as freedom, responsibility, rationality, and the mind-body relationship.

In essence, these theses and related concepts pertaining to neuroethics, the underpinnings of our moral or social behavior, and other related subjects could be collectively referred to as "great challenges." To elucidate its principles and place them in a more human and realistic perspective, it is necessary to invoke external concepts or frameworks, because as has been demonstrated in numerous studies, neuroethics is a discipline that lacks a comprehensive explanatory framework of its own.

Addressing these great challenges necessitates a multifaceted approach, encompassing diverse perspectives, scientific disciplines, and methodological frameworks. Iain McGilchrist's work (2021) defends that ancient spiritual truths, neuroscience, physics and the best philosophy, all lead us towards a world that makes sense as a whole. It is dangerous to drive them apart into their separate silos again. Science, according to McGilchrist, should give up its monopoly claims.

A multidisciplinary dialogue is imperative to facilitate the exchange of empirical, humanistic, and sapiential insights, grounded in a genuine pursuit of truth. It is crucial to recognize that such an endeavor is inherently limited in its scope and does

not claim to exhaust the subject matter. To continue thinking of body as a mechanistic being or of brain as the sufficient condition for our moral acts or for our social condition, shows an Achilles heel that neuroethics should explain with more detail.

References

Aristotle (1995, a). Nicomachean Ethics. In *The Complete Works of Aristotle*, vol. 2. Princeton: Princeton University Press, 1729-1867.

Aristotle (1995 b). On the Soul. In Barnes, J. (1995). *The complete works of Aristotle: the revised Oxford translation*. Vol. 1 (6th printing, with corrections). Princeton: Princeton University Press.

Aristotle (1995 c). Politics. In Barnes, J. (1995). *The complete works of Aristotle: the revised Oxford translation*. Vol. 2 (6th printing, with corrections). Princeton: Princeton University Press.

Benedict XVI (1995). *In the beginning…*: a Catholic understanding of the story of Creation and the fall. Grand Rapids: W.B. Eerdmans Pub. Co.

Cortina, A. (2013). Ética del discurso: ¿un marco filosófico para la neuroética? *Isegoría. Revista de Filosofía Moral y Política*, n. 48, enero-junio, 127-148.

Cortina, A. (2010). Neuroética: ¿Las bases cerebrales de una ética universal con relevancia política? *Isegoría. Revista de Filosofía Moral y Política*, n. 42, enero-junio, 129-148.

Cortina, A. and Conill, J. (2019). Bioética y neuroética. *Arbor*, 195 (792): a503. https://doi.org/10.3989/arbor.2019.792n2004

Churchland, P. (1986). *Neurophilosophy*. Toward a Unified Science of the Mind-Brain. Cambridge, Mass.: MIT Press.

Descartes, R. and Moriarty, Michael. (2015). *The passions of the soul and other late philosophical writings*. Oxford, United Kingdom; New York: University Press.

González, A.M. (2004). La dignidad de la persona, presupuesto de la investigación científica. Concepciones de la dignidad. In: *Biotecnología, dignidad y derecho: bases para un diálogo*, Ballesteros, J.; Aparisi, A. (eds.). Pamplona: EUNSA, 17-41.

Heidegger, M. (1977). *The question concerning technology, and other essays* (1st ed..). New York: Harper & Row.

Inciarte, F. (1979). Die Seele aus Begriffsanalytischer Sicht. In Seebab, H. (ed.). *Entstehung des Lebens. Studium generale Wintersemester 1979/1980*. Aschendorf, 47-70.

Jonas, H. (1984). *The imperative of responsibility*: foundations of an ethics for the technological age. Chicago: University of Chicago Press.

Kant, I. and Abbott, Thomas Kingsmill (1988). *Fundamental principles of the metaphysic of morals*. Amherst, N.Y.: Prometheus Books.

Kotowicz, Z. (2007). The strange case of Phineas Gage. *History of the Human Sciences*, *20*(1), 115-131.

MacIntyre, A. C. (1999). *Dependent rational animals*: why human beings need the virtues. Chicago, Ill.: Open Court.

MacIntyre, A. (2006). *The tasks of philosophy: Selected essays,* (Vol. 1). Cambridge, UK; New York: Cambridge University Press.

McGilchrist, I. (2021). *The matter with things*: our brains, our delusions, and the unmaking of the world. London: Perspective Press.

Pico della Mirandola, G., Borghesi, Francesco, Papio, Michael and Riva, Massimo (2016). *Oration on the dignity of man: a new translation and commentary*. Cambridge, UK, New York: Cambridge University Press.

Rizzolatti, G., and Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. *Current opinion in neurobiology*, 18(2), 179-184.

Roskies, A. (2002). Neuroethics for the new millenium, *Neuron*, 35, 21-23.

Spaemann, R. and Madigan, Arthur. (2000). *Happiness and benevolence*. South Bend: University of Notre Dame Press.

Spaemann, R. (1994). *The Inviolability of Human Life*. Philosophical Yearbook, (27), 69-87.

— 2 —

The belief in a just world: neuroscientific perspectives and legal implications

MURILO KARASINSKI
Pontifícia Universidade Católica do Paraná (Brazil)
k.murilo@pucpr.br

RENATO CESAR CARDOSO
Universidade Federal de Minas Gerais (Brazil)
renatoccardoso@hotmail.com

Introduction

The human brain doesn't cope well with unpredictability, finding it difficult to accept that something can happen for no reason other than chance. Taleb (2019) argues, for example, that the success of free trade is not based on rewards or incentives for those with the best techniques but rather on its ability to create an environment favorable to luck, especially in contexts marked by trial and error. Despite this, the idea of luck is often overshadowed by notions of merit and effort. As Botton (2021) suggests, this led to the development, particularly in the 19th and 20th centuries, especially in Europe and the United States, of the belief that there is a correlation between wealth and virtue. Thus, if the successful *deserve* their achievements, the other side of the coin implies that the unsuccessful must equally *deserve* their failure.

This difficulty in acknowledging unpredictability results in a tendency to attribute reasons to events, often interpreted as "fate". Simply labeling someone as unlucky does not satisfy a brain eager for explanations. As Burnett (2018) observes, the aversion to uncertainty leads to extreme conclusions, even when supported by weak reasoning. This helps explain why superstitions and conspiracy theories continue to thrive in the 21st century. It also clarifies the phenomenon of self-deception and the search for seemingly reasonable explanations for events – often after they occur – to impose order and causality on reality. Part of this is related to the availability heuristic,

which Kahneman (2012) defines as the process of judging frequency based on the 'ease with which occurrences come to mind. Essentially, the availability heuristic, like other judgment heuristics, substitutes one question for another.

According to Kahneman (2012), it is easier to overestimate the number of divorces among celebrities or the number of sex scandals involving politicians simply because such events receive greater media attention – for obvious reasons – than similar cases involving non-famous individuals. Likewise, airplane safety is often underestimated after high-profile crashes, leading to the false belief that cars are safer. What connects all these cases is that rare events attract disproportionate attention. As Kahneman (2012) explains, our expectations about the frequency of events are distorted by the prominence and emotional intensity of the messages we are exposed to. By distorting how we perceive the past, the availability heuristic also contributes to the creation of false narratives, generating, as Taleb (2019) states, an *illusion of understanding* – namely, the fantasy that everyone thinks they know what is happening in a world that is more complex (and random) than they realize.

This distortion in how we interpret reality not only fuels misconceptions about frequency and risk but also reinforces broader beliefs about the justice of the world. If striking and widely publicized events become reference points for assessing normality, it is natural that events associated with success or failure follow the same logic. People who become symbols of prosperity are remembered for their achievements, while individuals who face misfortune often have their trajectories reduced to aspects that seem to justify their condition. In this way, the availability heuristic reinforces the tendency to perceive the world as an orderly and morally structured place, where people receive what they deserve. Thus, the need to minimize chance and randomness in the world gives rise to the central belief explored in this chapter: the Belief in a Just World (BJW). Its foundations and repercussions will be presented below. In essence, the Belief in a Just World reflects the effort to attribute purpose to reality, assuming that actions hold meaning and that, ultimately, virtuous behavior will be rewarded, while improper conduct will be punished.

The belief that the world operates in a just and predictable manner is not merely a psychological phenomenon but also a deeply ingrained element in cultural and religious systems. Since antiquity, narratives have been created to explain the relationship between conduct and consequence, often linking suffering and reward to a higher moral order. One of the most emblematic examples illustrating the Belief in a Just World can be found in the narrative of the Book of Job in the Old Testament. The story recounts how Satan challenges God to test the faith of an upright, just, and God-fearing man who had seven sons, three daughters, vast herds, and many servants. To prove Job's loyalty, God allows (Job 1:12) him to endure a series of tragedies: his children are killed, his property is destroyed, and his body is afflicted with sores from the soles of his feet to the top of his head (Job 2:7). Job's friends suspect that he must have committed some hidden sin, believing that divine wrath has turned against him for a just reason. They insist that Job, in some way, must deserve the suffering he faces, despite his repeated declarations of innocence.

In philosophy, the idea that everything happens for a reason is associated with the concept of theodicy, a term introduced in the work *Essays on Theodicy: On the Goodness of God, the Freedom of Man, and the Origin of Evil* by Leibniz (2023),

originally published in 1710. Theodicy seeks to resolve Epicurus' paradox, which challenges three attributes of the Judeo-Christian God: omnipotence, omniscience, and omnibenevolence. If God is omnipotent and omniscient, He knows all evil and has the power to eliminate it but does not do so, which would compromise His omnibenevolence. If He is omnipotent and omnibenevolent, He would have the power and the desire to eradicate evil, but by not knowing its full extent, He would cease to be omniscient. Finally, if He is omniscient and omnibenevolent, He knows all evil and wishes to eliminate it, but the inability to do so would compromise His omnipotence, challenging the logic of the existence of a God who simultaneously possesses all three attributes.

Whether in religious or secular contexts, we invoke the problem of evil whenever we protest against something we perceive as unjust or believe should exist in a fairer manner. Similarly, we resort to a kind of theodicy when justifying that a given situation, despite its apparent flaws, represents "the best possible" within a broader context, as in the thought of Leibniz, later satirized in *Candide, or Optimism* by Voltaire (2012). This reasoning connects to the core of the Belief in a Just World, which leads to the idea that people who have faced terrible misfortunes must, in some way, have done something to deserve them, as seen in the suspicions of Job's friends. This chapter aims to investigate this belief from a neuroscientific perspective, exploring some of its legal implications.

1. Belief in a just world: from the construction of the concept to victim blaming

The conceptual foundations of the Belief in a Just World can be traced to the work of Melvin Lerner (1980), who coined the term based on his observations while working in a hospital. Lerner (1980) noticed that a significant number of healthcare professionals attributed the blame for the condition of mentally ill patients to the patients themselves, reinforcing the idea that the world is inherently just. These findings extended to other contexts as well. For example, in the early 1960s, at the University of Kentucky, Lerner (1980) conducted an experiment designed to investigate how people adjust their perceptions to justify outcomes perceived as arbitrary. The study involved groups of approximately five female university students, who participated as observers in a controlled experimental setting. The experiment was structured as follows: the participants observed two individuals performing an intellectual task of "word formation". Both demonstrated similar performance levels, with no significant differences in the quality or quantity of their results. However, the distribution of rewards was entirely random, unrelated to merit or performance. One of the individuals received a reward, while the other received nothing.

After witnessing the task and the allocation of rewards, the participants were asked to evaluate the performance of the two individuals. The objective was to explore how they would interpret the scenario considering the evident injustice. The results revealed that the observers frequently reinterpreted the situation to justify the favoritism. Many began to believe that the rewarded worker was more competent or deserving than the other, even when confronted with objective evidence that their performance was identical.

In another case, Lerner (1980) investigated how people reacted when witnessing the suffering of a perceived innocent victim. During the experiment, participants watched a recording in which a woman appeared to be taking part in a learning test. In this simulated scenario, the woman received electric shocks as punishment for incorrect answers. The victim, an individual trained by the experimenters, exhibited intense signs of pain, such as groans and abrupt physical reactions, creating a realistic impression of suffering. After observing the scenario, participants were asked to evaluate the victim, considering attributes such as likability, competence, and responsibility for her suffering. The results revealed an intriguing pattern: rather than expressing empathy or outrage at the injustice suffered by the victim, many observers devalued her. Comments such as "weak," "unprepared," or "incompetent" frequently appeared in descriptions of the woman (victim).

Another interesting study conducted by Lerner (1980) exposed participants to two distinct situations involving a new victim. In the first scenario, the victim was presented as someone who accepted suffering as a moral or academic necessity. She appeared to endure the situation with resignation, indicating that she believed she was contributing to a greater good. In the second scenario, the victim's suffering was portrayed as purposeless and unjustified, highlighting an arbitrary and meaningless condition. Participants were then asked to describe their impressions of the victim and her situation. In the case of the "martyr", who faced suffering with an apparent altruistic or academic motivation, observers tended to show more respect or sympathy. However, even in this context, it was not uncommon for some participants to devalue the victim, perceiving her as naive or overly idealistic. In contrast, in the case of the victim who suffered without apparent purpose, participants responded more harshly and with less compassion. The victim was described as responsible for her own condition, receiving judgments that sought to attribute blame or personal flaws to justify her suffering.

Unfortunately, the Belief in a Just World is not limited to controlled experiments in university settings. A study conducted by the Brazilian Public Security Forum (2016) revealed concerning data: one in three Brazilians believes that, in cases of rape, the victim is to blame. Among men, this belief is even more prevalent: 42% agree with the statement that "women who respect themselves are not raped". This victim-blaming logic is not exclusive to men. Among women – who are the primary victims of this crime – 32% also agree with this idea. Additionally, 30% of men believe that a woman who wears provocative clothing has no right to complain if she is raped.

In another context, analyzing teenagers' perceptions of the causes of bullying in Sweden, Thornberg and Knutsen (2011) found that many young people explain bullying because of the victim's own characteristics or behaviors – a perspective that reflects the idea that people get what they deserve. Approximately 42% of adolescents attributed blame to the victim, justifying bullying based on supposed deviations or perceived differences, such as physical, social, or behavioral traits that made the victim a target. This suggests that young people internalize the notion that victims are responsible for their suffering, whether due to failing to conform to social norms or exhibiting behaviors that attract hostility. This narrative also provides moral justifi-

cation for aggressors, who, by blaming the victim, minimize their own responsibility and reinforce the idea that their actions are inevitable or justified.

The notion that the world is unjust, where bad things happen to good people without any apparent explanation – such as traffic accidents or the diagnosis of malignant tumors – is uncomfortable and distressing. Nevertheless, this is an everyday reality. Therefore, to gain a deeper understanding of the Belief in a Just World, it is essential to explore how the phenomenon of cognitive dissonance operates, a topic that will be addressed in the next section of this chapter.

2. The role of cognitive dissonance

According to Dobelli (2014), the fable by the Greek poet Aesop illustrates one of the most famous thinking errors: A fox, drawn to a vine, fixed its gaze on robust, ripe, purple grapes. Resting its front paws on the trunk, it stretched its neck and tried to reach the fruit, but they were out of its grasp. Frustrated, it made another attempt, this time jumping, but again, it only touched the air. On the third try, it gathered all its strength for an even bigger leap but ended up falling on its back. The grapes remained untouched, still hanging from the vine. Faced with failure, the fox wrinkled its nose and said, "They are not ripe enough for me anyway; I don't like sour grapes". With its head held high, it returned to the forest, proud.

In this case, the fox's pretentious attitude and the frustrating outcome of its action are in conflict. To deal with this uncomfortable discrepancy, it could choose among three alternatives: (a) keep trying to reach the grapes in some way; (b) acknowledge its inability to reach them; or (c) reinterpret the situation. This last alternative, according to Dobelli (2014), exemplified in the fox's questionable justification about the taste of the grapes, is what characterizes *cognitive dissonance*.

To test cognitive dissonance, a fundamental question is what happens to people with unwavering faith when their convictions are challenged by reality. In the book *When Prophecy Fails* (1956), psychologists Leon Festinger, a pioneer in cognitive dissonance studies, Henry Riecken, and Stanley Schachter narrated the story of Dorothy Martin, a woman who claimed to communicate with a group of aliens called the Guardians. According to Dorothy, these Guardians had warned that the world would be destroyed by a great flood on December 21, 1954. However, she and her followers would be saved by a spaceship that would land in her backyard a few days before the cataclysm.

Preparing for the rescue, Dorothy's followers made drastic decisions, such as quitting their jobs, abandoning their families, and donating all their belongings. With full confidence in the prophecy, they gathered at Dorothy's house in Chicago to await the arrival of the spaceship. Initially, Dorothy announced that the ship would land at 4:00 PM on December 17, marking the beginning of their salvation.

But the aliens did not come. At midnight on the 17th, Dorothy claimed to have received a new message: the ship was on its way. However, the leader then communicated that the aliens had provided a new date – midnight on December 21, just before the apocalypse. Once again, the Guardians gathered in Dorothy's living room to wait. For Festinger *et al*. (1956), it was evident that the prophecy would not come

true. Still, a group of people remained committed to the prophecy, treating it as an unquestionable reality.

As the clock showed only one minute remaining until the arrival of the flying saucer, the Guardians remained still, showing no visible reaction. Midnight passed, and still, nothing happened. Despite this, no one in the room moved or spoke; their faces were frozen, expressionless. Hours passed until, little by little, they accepted that no visitors from outer space would come to rescue them. Did this "disconfirmation" shake their beliefs? No. At 4:45 AM that morning, Dorothy announced another message. This time, she stated that, due to the unwavering faith of the Guardians, God had canceled the destruction of the world.

The story of Dorothy Martin illustrates how cognitive dissonance can lead individuals to reinterpret reality to preserve their prior beliefs, even in the face of contradictory evidence. This same mechanism is present in the Belief in a Just World. When confronted with situations that challenge the idea that people get what they deserve, many individuals prefer to reframe their interpretation of the facts rather than accept that the world can be chaotic and unpredictable. As a result, they justify others' suffering by attributing it to some moral or personal cause, thereby reducing the dissonance between their worldview and reality. This process, though often unconscious, strengthens the tendency to blame victims and reinforces punitive beliefs that impact social and legal judgments, as will be discussed later in this chapter.

For Festinger *et al*. (1956), we have all experienced the futility of trying to change a strong conviction, especially when the person holding it has some kind of investment in the belief. We are familiar with the variety of ingenious defenses people use to protect their convictions, keeping them intact even in the face of the most devastating challenges. However, human ingenuity goes beyond simply protecting a belief. According to Festinger *et al*. (1956), suppose an individual believes in something with all their heart; suppose, furthermore, that they have committed to this belief, taking irreversible actions because of it; and finally, suppose they are confronted with unequivocal and undeniable evidence that their belief is wrong – what will happen? In many cases, the individual does not emerge shaken but even more convinced of the truth of their beliefs than ever before. In fact, they may even display newfound zeal in trying to convince and convert others to their view. In other words, the more someone invests in a set of beliefs: the greater the sacrifice made in their name, the more resistant that person becomes to contrary evidence. Instead of giving up, they redouble their bet.

Unfortunately, as Pinker (2013) explains, refuting someone's belief often threatens their dignity, status, and sense of power. When that belief is based solely on faith, it becomes fragile. According to Pinker (2013), no one feels threatened by believing that stones fall downward, as any mentally sound person can verify this fact with their own eyes. The same, however, does not apply to beliefs such as the idea that babies are born with original sin, that God exists as three persons, or that Ali is the second most divinely inspired man after Muhammad. When individuals structure their lives around such beliefs and discover that others seem to live perfectly well without them – or worse, credibly refute them – Pinker (2013) argues that they feel threatened by the possibility of being perceived as naive or mistaken. Since it is impossible to defend a belief based solely on faith by convincing skeptics of its truth,

disbelief often provokes reactions of anger. Many believers, when confronted with a challenge to what gives meaning to their lives, choose to try to silence or eliminate the voices that threaten their convictions.

This type of behavior is uniquely exemplified in George Orwell's *1984* (1949), where the control of reality was referred to as doublethink. The three slogans of the Ministry of Truth – "war is peace," "freedom is slavery," "ignorance is strength" – reinforced the ability to simultaneously hold two contradictory beliefs and accept both as true. This process had to be conscious enough to be executed with precision, yet unconscious enough to prevent the contradiction from triggering a sense of falsehood and, consequently, a feeling of guilt. Cognitive dissonance becomes evident in this reflection by the main character:

> Winston sank his arms to his sides and slowly refilled his lungs with air. His mind slid away into the labyrinthine world of doublethink. To know and not to know, to be conscious of complete truthfulness while telling carefully constructed lies, to hold simultaneously two opinions which cancelled out, knowing them to be contradictory and believing in both of them, to use logic against logic, to repudiate morality while laying claim to it, to believe that democracy was impossible and that the Party was the guardian of democracy, to forget whatever it was necessary to forget, then to draw it back into memory again at the moment when it was needed, and then promptly to forget it again: and above all, to apply the same process to the process itself. That was the ultimate subtlety: consciously to induce unconsciousness, and then, once again, to become unconscious of the act of hypnosis you had just performed. Even to understand the word 'doublethink' involved the use of doublethink. (ORWELL, 1949)

According to Wegner (2002), the most intense dissonance occurs when someone performs an action that contradicts a previously established attitude or desire. Buying a house with an old and deteriorated boiler, for example, can generate cognitive dissonance. Despite knowing the problem they will face with the heating system, the person decides to proceed with the purchase. Dissonance suggests that, in such situations, individuals adjust their attitude to align it with their behavior. The attitude (in this case, neutrality toward the house) is easier to change since it exists in private thoughts, whereas the act of purchasing the house (with its problematic heating system) is public and more difficult to reverse. Interestingly, according to Wegner (2002), the person ends up liking the house precisely because of the problem with the boiler.

As Burnett (2018) explains, when we witness someone suffering something terrible and undeserved, dissonance arises: we believe the world is just, but the event contradicts this view. Faced with this conflict, the brain seeks to resolve it in two ways: (a) accepting that the world can be cruel and meaningless, or (b) concluding that the victim, in some way, brought misfortune upon themselves. Although the second alternative is harsher, it preserves the (misguided) belief that the world is essentially good and welcoming. As a result, we tend to blame victims for their tragedies, shifting the responsibility for suffering onto them.

3. Neuroscientific findings and legal implications

From a neuroscientific perspective, Denke *et al*. (2014) explored the neural basis of individual differences in the Belief in a Just World when observing norm violations. Using fMRI, research participants analyzed scenarios involving behaviors that either violated or confirmed norms and provided moral evaluations of the described actions. The main hypothesis was that this belief would be related to neural activity

in regions associated with the perception of justice and morality. According to Denke *et al*. (2014), activity in the insula and somatosensory cortices showed a significant correlation with the Belief in a Just World. This finding suggests that these regions not only process social norm violations but also reflect individual differences in how people perceive and react to such violations, depending on their beliefs. While the insula is known for integrating negative emotions and moral perceptions, the somatosensory cortices appear to contribute to the somatic dimension of empathy, meaning the ability to feel other people's experiences as one's own. Denke *et al*. (2014) argue that the relationship between Belief in a Just World and neural activity can be explained by the hypothesis that individuals with a stronger belief in this idea tend to exhibit more intense responses when witnessing injustices, which is reflected in greater activation of the insula.

Regarding brain function, Hafer *et al*. (2022) examined how Belief in a Just World relates to neural processing of information about deservedness, using event-related potentials (ERPs). Participants read stories where outcomes were deserved, undeserved, or neutral, while their ERPs were recorded. The goal was to identify whether individual differences in Belief in a Just World influenced the rapid processing of these types of information. According to Hafer *et al*. (2022), the results revealed that Belief in a Just World is associated with automatic processing of deservedness-related information. Participants with a stronger belief in this idea displayed distinct neural responses, particularly in the early stages of processing (less than 300 ms after the stimulus) when confronted with undeserved outcomes compared to deserved or neutral ones. This pattern suggests that, for these individuals, outcomes that contradict their expectations of social justice attract greater attention, indicating higher salience for these events. Thus, Hafer *et al*. (2022) argue that Belief in a Just World directly influences automatic attention processes at very early stages of the central nervous system, occurring before individuals can consciously deliberate on the stimuli. More specifically, the results showed that, in people with a stronger Belief in a Just World, perceived "undeserved" outcomes generated distinct neural responses in extremely fast time windows, such as 96 ms after the stimulus. This timeframe is associated with automatic attention processes, indicating that their brains not only detect but also react in a way consistent with the expectation that the world is orderly and predictable.

A meta-analysis conducted by Nudelman (2013) investigated the relationship between BJW and personality traits within the five-factor model. The study synthesized data from twelve independent studies, providing a comprehensive view of the correlations between BJW and neuroticism, extraversion, and agreeableness. The results indicated that BJW negatively correlates with neuroticism and positively correlates with extraversion and agreeableness. According to Nudelman (2013), the negative correlation between BJW and neuroticism suggests that Belief in a Just World may function as a psychological resource, promoting emotional stability and reducing anxiety and depression. This relationship can be explained by BJW's role as a coping mechanism, helping individuals deal with adversity by interpreting events in a more predictable and orderly manner. However, causality cannot be established, as it is possible that individuals with greater emotional stability develop a stronger BJW over time.

According to Nudelman (2013), the positive association between BJW and extraversion suggests that individuals who perceive the world as a just place tend to be more sociable and confident in their interpersonal interactions. The belief that people get what they deserve may reduce perceived risk and, consequently, encourage more expansive and proactive behaviors. One possible explanation for this relationship is that BJW provides a basis for psychological security, allowing greater openness to social contact. Another hypothesis suggests that extroverted individuals, by interacting with a socially responsive environment, reinforce their belief in the world's justice. Nudelman (2013) also confirmed the relationship between BJW and agreeableness, possibly due to interpersonal trust and respect for social order, traits common to both characteristics.

For Butler and Moran (2007), Belief in a Just World directly influences legal decisions, particularly in the context of capital punishment trials. Their research demonstrated that jurors who support the death penalty tend to have a higher BJW, which may compromise the impartiality of the process, affecting the final sentencing decision. According to Butler and Moran (2007), BJW is associated with a juror profile that believes in the inherent justice of the legal system and, consequently, shows a greater tendency to accept prosecution arguments and trust prosecutors' testimony. This phenomenon is related to a form of authoritarianism that emphasizes submission to norms and state authority. Individuals with high scores in this trait are more inclined to recommend the death penalty and less receptive to arguments emphasizing rehabilitation or the social context of the crime.

According to Butler and Moran (2007), the exclusion of jurors who do not qualify for a capital punishment trial results in a more homogeneous jury pool that is predisposed to conviction. This systematic selection favors a punitive view of the law, reducing the diversity of perspectives that could lead to a more balanced trial. Additionally, high BJW can lead jurors to disregard systemic errors, reinforcing their belief in the infallibility of the criminal process. Butler and Moran (2007) argue that the impact of BJW on legal decisions raises questions about the fairness of the jury system in capital punishment cases. The tendency of jurors to justify severe punishments based on the idea that the world operates justly compromises their ability to evaluate the defendant's individual circumstances.

For Scherr and Franks (2015), Belief in a Just World has a significant impact on how innocent and guilty suspects make decisions during the pre-interrogation phase. Their meta-analysis demonstrated that BJW influences the decision to waive or invoke interrogation rights, a determining factor in the legal unfolding of cases. Innocent individuals, trusting that justice will prevail, are more likely to waive their rights, making them vulnerable to pressures that can lead to wrongful convictions.

According to Scherr and Franks (2015), BJW affects innocent and guilty suspects differently. Innocents with a strong belief in the world's justice show less susceptibility to social pressures encouraging them to waive their rights, as they assume that the truth will protect them. In contrast, innocents with low BJW are more susceptible to external influences, leading them to decisions that may compromise their defense. This dynamic reinforces the vulnerability of suspects who, trusting in the system's impartiality, ultimately undermine their legal position.

Among guilty suspects, BJW operates inversely. Guilty individuals with high BJW experience higher stress levels when accused, as they believe their transgressions will inevitably lead to punishment. This emotional state makes them more vulnerable to social influence strategies used to obtain confessions. On the other hand, guilty individuals with low BJW show less emotional impact when accused and greater resistance to coercion tactics, possibly because they perceive the system as flawed or manipulable.

Scherr and Franks (2015) argue that the legal implications of these findings are significant. Innocents' tendency to overtrust justice can lead to premature waivers of their rights, reducing their ability to defend themselves effectively. From a procedural perspective, this behavior can be exploited by interrogation techniques aimed at extracting self-incriminating statements. At the same time, the fact that guilty individuals with high BJW are more susceptible to stress and social influence suggests that the legal system itself may reinforce inequalities in the administration of justice, making certain individuals more likely to accept plea deals or confessions without resistance. According to Scherr and Franks (2015), the implementation of standardized protocols and awareness of BJW's impacts could help mitigate judicial errors and protect individuals from wrongful convictions.

Conclusions

The Belief in a Just World structures ways of interpreting events and judging behaviors, shaping perceptions of merit and responsibility. Assigning meaning to reality allows for organizing expectations and providing a sense of predictability, but it also conditions individual and collective responses to situations perceived as unjust. Neuroscience has explored how this belief manifests in the brain, identifying correlations between cognitive processes and the tendency to justify occurrences based on the idea of immanent justice.

In the legal field, the Belief in a Just World influences decisions regarding culpability and sentencing, shaping how individuals assess mitigating or aggravating circumstances. The way suspects, victims, and adjudicators interpret reality is permeated by this belief structure, which can compromise the impartiality of evaluations and reinforce biases that impact decision-making processes.

The studies analyzed indicate that this belief influences cognition at early processing levels, suggesting that its foundations are not merely social but also neurobiological. The tendency to perceive the world as orderly and governed by principles of reward and punishment is not limited to cultural or philosophical constructions but emerges as a mechanism that guides social interactions and moral perceptions.

The debate on the Belief in a Just World remains open. The relationships between cognition, morality, and justice continue to be a subject of investigation, especially considering social transformations that challenge traditional models of merit and responsibility. The connections between brain processes and normative structures suggest that further research could deepen the understanding of how these beliefs are established and evolve throughout life.

Job's friends, unable to conceive that a just person could suffer without cause, insisted that he must have committed some hidden wrongdoing to deserve such suf-

fering. This perspective reflects the tendency to associate misfortune with individual blame, a bias that neuroscience has helped to elucidate. By identifying the cognitive and emotional mechanisms that sustain this belief, neuroscientific research paves the way for a better understanding of how we judge others and, perhaps, for mitigating the impact of this tendency, especially in the legal field.

References

ATLAS DA VIOLÊNCIA. (2016). São Paulo: *Fórum Brasileiro de Segurança Pública*; Rio de Janeiro: IPEA, mar. 2016. ISSN 2764-0361

Butler, B., Moran, G. (2007). The impact of death qualification, Belief in a Just World, legal authoritarianism, and locus of control on venirepersons' evaluations of aggravating and mitigating circumstances in capital trials. *Behavioral Sciences and the Law*, v. 25, n. 1, p. 57-68, 2007. DOI: https://onlinelibrary.wiley.com/doi/10.1002/bsl.734.

Botton, A. (2021). *Desejo de status*. Porto Alegre: L&PM.

Burnett, D. (2018). *O cérebro que não sabia de nada*: o que a neurociência explica sobre o misterioso, inquieto e totalmente falível cérebro humano. São Paulo: Planeta do Brasil.

Denke, C., Rotte, M., Heinze, H-J., Schaefer, M. (2014). Belief in a Just World is associated with activity in insula and somatosensory cortices as a response to the perception of norm violations. *Social Neuroscience*, v. 9, n. 5, p. 514–521, 2014. DOI: https://doi.org/10.1080/17470919.2014.922493.

Dobelli, R. (2014). *A arte de pensar claramente*: como evitar as armadilhas do pensamento e tomar decisões de forma mais eficaz. Rio de Janeiro: Objetiva.

Festinger, L., Riecken, H. W, Schachter, S. (1956). *When prophecy fails.* University of Minnesota Press. https://doi.org/10.1037/10030-000

Hafer, C. L., Weissflog, M., Drolet, C. E. Segalowitz, S. J. (2022). The relation between Belief in a Just World and early processing of deserved and undeserved outcomes: An ERP study. *Social Neuroscience*, v. 17, n. 2, p. 95–116, 2022. DOI: https://doi.org/10.1080/17470919.2022.2038262.

Kahneman, D. (2012). *Rápido e devagar: duas formas de pensar*. Rio de Janeiro: Objetiva.

Leibniz, G. W. (2023). *Ensaios de teodiceia*: sobre a bondade de Deus, a liberdade do homem e a origem do mal. São Paulo: Editora WMF Martins Fontes.

Lerner, M. (1980). *The Belief in a Just World*. New York: Springer, 1980.

Nudelman, G. (2013). The Belief in a Just World and Personality: A Meta-analysis. *Social Justice Research*, v. 26, n. 2, p. 105-119, 2013. DOI: https://doi.org/10.1007/s11211-013-0178-y

Orwell, G. (1949). *1984*. London: Secker & Warburg.

Pinker, S. (2013). *Os anjos bons da nossa natureza*: Por que a violência diminui. São Paulo: Companhia das Letras.

Scherr, K. C., Franks, A. S. (2015). The world is not fair: an examination of innocent and guilty suspects' waiver decisions. *Law and Human Behavior*, v. 39, n. 2, p. 142–151, 2015. DOI: http://dx.doi.org/10.1037/lhb0000121.

Taleb, N. (2019). *A lógica do Cisne Negro*: o impacto do altamente improvável. Rio de Janeiro: BestBusiness.

Thornberg, R., Knutsen, S. (2011). *Teenagers' Explanations of Bullying*. Child Youth Care Forum 40, 177–192. DOI: https://link.springer.com/article/10.1007/s10566-010-9129-z

Voltaire (2012). *Cândido, ou o otimismo*. Porto Alegre: L&PM.

Wegner, D. (2002). *The illusion of conscious will*. Cambridge: MIT Press.

— 3 —

Paradoxes of transhumanism: impacts on equality, freedom and human dignity

JOSÉ OCTÁVIO DE CASTRO MELO
Universidade Estadual do Piauí (Brazil)
joseoctavio@edu.unifor.br

Introduction

The limitations imposed by nature reflect humanity's desire to transcend, marking human history as a constant pursuit of immortality and mental and physical enhancement. With advances in science and technology, this aspiration has become tangible through tools such as genetic editing, bionic prosthetics, nanotechnology, and artificial intelligence. These innovations not only enable interventions in human DNA and the recovery of lost functions but also significantly increase life expectancy. However, they raise ethical and social dilemmas that challenge the principles of freedom, equality, and dignity.

Genetic editing enables the correction of hereditary diseases and the development of personalized treatments, whereas bionic prosthetics and neural implants restore functions in individuals with severe disabilities. Nevertheless, the excessive costs of these technologies restrict accessibility to an economic elite, exacerbating social inequalities. This scenario highlights the risk of more profound biological and social divisions, with implications that extend beyond the field of health.

The boundaries of human existence raise philosophical and practical questions about the pursuit of immortality and physical and mental enhancement. While appealing, the indefinite extension of life could strain natural resources, intensify inequalities, and devalue the human experience, which is inherently defined by its finitude. In this context, it is essential to establish ethical regulations that promote the equitable distribution of technological benefits, preserving fundamental values such as freedom, equality, and human dignity.

Transhumanism, with its promise of human enhancement through advanced technologies, exposes the paradox between innovation and social justice. The inaccessibility of technologies such as genetic editing and bionic prosthetics reinforces inequalities, creating the risk of "biological castes" where only an elite has access to advancements that enhance longevity and human performance.

This article aims to analyze the impacts of transhumanism on equality, freedom, and human dignity in light of advanced human enhancement technologies. The specific objectives are a) to identify the paradox of equality in access to human enhancement technologies; b) to evaluate individual freedom of choice in the face of transhumanist technology; and c) to understand human dignity in the context of neurotechnological advancements.

The research question posed is: to what extent can transhumanism impact the principles of equality, freedom, and human dignity, considering the challenges of accessibility and the risks of exacerbating social inequalities in a context of rapid technological advancement?

Reports from international organizations, such as the United Nations (UN) and the Organization for Economic Co-operation and Development (OECD), emphasize the urgency of addressing these disparities. The OECD warns of the dangers of a biotech market that exclusively helps the wealthy, exacerbating social inequalities. The research will employ a qualitative approach based on bibliographic review and document analysis.

The study is expected to identify evidence that transhumanist technologies, while promising, may exacerbate social inequalities by being predominantly accessible to economic elites. The research aims to propose ethical and regulatory guidelines to ensure that the benefits of these technologies are widely distributed. Potential solutions include implementing government subsidies for universal access, fostering public-private partnerships to reduce costs, and establishing international regulatory frameworks that promote equity.

1. The paradox of equality: a privilege for the few or innovation for all?

Transhumanism carries the promise of democratizing opportunities and improving quality of life, yet its advanced technologies often exacerbate social and economic inequalities. Innovations such as genetic editing and neural implants, due to their excessive costs, remain accessible only to a privileged elite, leaving much of the population excluded from these advancements. This raises the question: are these innovations truly fostering inclusion, or are they deepening existing social divides?

Greek mythology offers clear metaphors for the dilemmas of transhumanism. Prometheus, who stole fire from the gods to give it to humanity, symbolizes the transformative power of technological innovations. However, much like the "fire" in the myth, modern technological advancements are not distributed equitably, benefiting only an elite while excluding the majority (Aeschylus, 2013). Similarly, Icarus embodies the risks of unchecked ambition and the lack of ethical boundaries. By ignoring *Daedalus*' warnings and flying too close to the sun, he met his downfall (Ovid, 2008).

Plato proposed the idea of ethical and intellectual enhancement as a means to achieve justice and wisdom. In *The Republic*, he suggested that the soul could be refined through reason, transcending basic instincts, and contributing to a more just society (Plato, 2006). These ideas laid the foundation for the pursuit of not only physical but also moral enhancement.

The Renaissance marked a renewed interest in human potential. Leonardo da Vinci's *Vitruvian Man* exemplifies the Renaissance ideal of harmony between body and mind, inspired by mathematical proportions and portraying the human body as a work of art worthy of scientific study and improvement (Kemp, 2019). In literature, *Frankenstein*[1] addressed the dangers of overstepping ethical boundaries in the pursuit of perfection. Victor Frankenstein's creation of artificial life raises questions about the implications of "playing God" and the responsibilities accompanying scientific advancements (Shelley, 1818).

The discovery of penicillin by Alexander Fleming in 1928 revolutionized medicine by saving millions of lives. More recently, technologies such as the CRISPR[2] genetic editing tool have promised precise alterations to human DNA, paving the way for the elimination of hereditary diseases (Doudna & Sternberg, 2017). While technologies like genetic editing and bionic prosthetics offer significant advancements, their excessive costs often reinforce inequalities. Tools like CRISPR, which can correct genetic disorders, remain inaccessible to most, raising concerns about the emergence of a "genetically enhanced race" with significant biological advantages (Fukuyama, 2002).

Genetic enhancement risks compromising individual autonomy by turning people into designed artifacts, fundamentally unequal in essence (Habermas, 2003). Similarly, neural implants like Neuralink[3] promise to enhance cognitive capacities but remain prohibitively expensive, limiting their accessibility to economic elites (Musk, 2020).

In the field of bionics, nerve-controlled prosthetics and cochlear implants have transformed the lives of people with physical disabilities. Modernity is marked by scientific and technological advances that have profoundly altered the human condition. Innovations such as modern medicine, artificial intelligence (AI), and human-machine integration have expanded human capabilities, overcoming natural limitations. Life expectancy has increased, fatal diseases have become treatable, and human-machine integration is now a reality.

Beyond socioeconomic disparities, these technologies may also compromise individual freedom. The United Nations emphasizes that autonomy and freedom of choice are fundamental principles that may be threatened by social pressures to adopt

[1] This work has become a landmark in the discussion of dilemmas associated with human transformation. With the advent of the Industrial Revolution and, later, biotechnological advancements, human enhancement moved beyond the realm of ideas to become a concrete objective.

[2] The CRISPR-Cas9 system (Clustered Regularly Interspaced Short Palindromic Repeats) is a revolutionary gene-editing technique that allows for precise and efficient modification of DNA sequences. Originally identified as an adaptive defense mechanism in bacteria against invading viruses, CRISPR-Cas9 was adapted for biotechnology, enabling gene editing in various organisms. Its applications range from correcting genetic mutations to enhancing agricultural traits, making it a fundamental tool in contemporary genetic research (Hupffer & Berwig, 2020).

[3] Neuralink, a neural implant technology, is designed to enhance cognitive capabilities but remains inaccessible to most due to its high initial cost, benefiting only a privileged minority (Musk, 2020).

enhancement technologies (UN, 2023). In competitive labor markets, implicit coercion to use neural implants or physical enhancements could undermine individual autonomy.

Internationally, the Oviedo Convention, promoted by the Council of Europe, prohibits non-therapeutic genetic interventions on the human genome, establishing an ethical regulatory framework (Council of Europe, 1997). Radical interventions, such as genetic modifications in embryos, could have disastrous consequences for humanity, including the loss of individual autonomy and deepened social inequalities (Habermas, 2003).

Inequalities are also evident in the practical application of these technologies. In sports, the use of advanced prosthetics raises debates about unfair advantages, while in education, tools such as virtual reality widen the gap between developed and developing countries (Morgan, 2018; World Bank, 2021). AI algorithms, when poorly trained, perpetuate historical biases, harming already marginalized groups (O'Neil, 2016).

Kurzweil celebrates advances that promise to overcome biological limitations, arguing that technologies such as nanotechnology and AI can enable humanity to transcend physical and mental constraints, approaching a technological singularity (Kurzweil, 2005). Transhumanism extends humanity's desire for self-improvement, fostering advancements in health, cognition, and longevity (Bostrom, 2003). Despite these advances, human enhancement raises ethical dilemmas. Habermas criticizes interventions that instrumentalize genetics, warning of the risks of dehumanization and inequality (Habermas, 2003). Sandel also questions the pursuit of perfection, asserting that it may undermine the acceptance of human imperfections, a fundamental basis for equality (Sandel, 2007).

Transhumanism, a movement that advocates the use of advanced technologies to enhance human capabilities, raises significant ethical questions, particularly regarding the principles of equality, freedom, and human dignity. Both the United Nations (UN) and the Organization for Economic Co-operation and Development (OECD) have expressed concerns about the implications of these technological innovations.

The UN, through the Universal Declaration of Human Rights, states that "all human beings are born free and equal in dignity and rights" (UN, 1948). However, unequal access to transhumanist technologies may exacerbate existing social disparities, contradicting this fundamental principle. For example, if only individuals from higher socioeconomic classes have access to enhancement technologies, inequalities may deepen, creating a more divided society.

The OECD warns of the risks of a biotech market that exclusively benefits the wealthy, potentially amplifying social inequalities (OECD, 2024). This evidence underscores the need for public policies to ensure equitable access to technological innovations, preventing the formation of an "enhanced elite" and ensuring that the benefits of transhumanism are shared across society.

Regarding freedom, the UN highlights the importance of individual autonomy and protection against manipulations that may compromise freedom of choice (UN, 2023). While transhumanist technologies offer potential enhancements, they may also introduce social pressures for adoption, limiting true individual freedom. For

instance, in highly competitive work environments, individuals may face implicit pressure to adopt certain enhancement technologies to maintain employability, undermining their autonomy.

In the United States, the Food and Drug Administration (FDA) regulates genetic therapies, while in the European Union, the European Medicines Agency (EMA) performs a similar role. Both organizations permit therapeutic research but prohibit heritable genetic modifications. In China, where regulations are less stringent, the case of scientist He Jiankui, who genetically altered embryos in 2018, resulted in legal condemnation and reinforced existing norms (Cave, 2020).

Fukuyama criticizes transhumanism for undermining fundamental human equality. He argues that genetic modification and other enhancements could create a caste-like society, where genetically enhanced individuals enjoy better opportunities and quality of life (Fukuyama, 2002). This biological dystopia is amplified by the fact that enhancement choices often fall on parents, introducing biases and inequalities from birth. Habermas similarly warns that genetic enhancement could compromise individual autonomy, transforming people into designed artifacts and making them fundamentally unequal (Habermas, 2003). He argues that human equality is rooted in natural unpredictability and diversity, both of which are threatened by enhancement technologies.

The paradox between enhancement and equality becomes evident in the context of limited resources. Technologies like bionic prosthetics and genetic editing, initially costly and exclusive, primarily benefit developed countries and privileged populations, exacerbating the exclusion of those who need them most, such as individuals with disabilities or rare genetic conditions in poorer nations. While promising liberation, these innovations also pose risks of inequality. Ensuring that progress benefits everyone requires collective efforts to preserve the principles of equality. As in the myth of Prometheus, knowledge must be shared equitably, illuminating humanity as a whole.

There is also the risk that technological advancements could become tools of exclusion rather than inclusion. An example is the use of AI algorithms in job recruitment processes, which have already demonstrated the potential to reproduce existing biases in historical data, such as racial and gender discrimination (O'Neil, 2016). The same technology that promises enhancement can, paradoxically, perpetuate inequalities.

The integration between humans and machines raises questions about identity and freedom. Neural implants, such as Neuralink proposed by Elon Musk, promise to enhance cognitive capabilities but also carry risks of external control or mental manipulation (Musk, 2020). Who will ensure that these devices are used ethically and safely?

2. The paradox of freedom: expanded autonomy or covert control?

Human freedom, widely promoted by modern technologies, also faces significant threats. Neural implants, which promise to enhance cognition or treat disorders, may introduce forms of external control and manipulation, representing risks of abuse by governments or corporations. Social pressure to adopt these advancements

can transform personal choices into obligations, limiting individual autonomy. Conversely, refusing to adopt certain technologies may lead to social or economic exclusion, raising the question of whether these innovations truly expand freedom or create new forms of dependency and inequality.

Transhumanism posits that technologies such as bionic prosthetics, brain-machine interfaces, and genetic editing can liberate humans from biological limitations. The integration of humans and machines is expected to enable unprecedented cognitive freedom, allowing individuals to access and process information at superior levels (Kurzweil, 2005). Technologies like Neuralink promise to enhance cognitive capacities and treat conditions such as paralysis, restoring autonomy for millions of individuals (Musk, 2020).

Additionally, CRISPR genetic editing offers the possibility of eradicating hereditary diseases, freeing individuals, and families from the burden of debilitating genetic conditions (Doudna & Sternberg, 2017). This technology could ensure that individuals are born with a "fair starting point," reducing health inequalities and promoting greater autonomy.

Despite these promises, the growing integration of humans and technology may limit freedom in unprecedented ways. The use of genetic enhancement technologies risks compromising autonomy by transforming individuals into designed products, whose capabilities are predetermined by external choices. Human freedom relies on the inherent unpredictability of nature, which would be lost in a world of planned enhancements (Habermas, 2003).

The dependency on technological devices like neural implants, while theoretically increasing cognitive autonomy, also introduces risks of external control, such as manipulation of thoughts or actions by governmental or corporate entities. The dangers posed by opaque algorithms in areas such as education, work, and healthcare highlight how automated decisions can restrict freedom rather than expand it (O'Neil, 2016).

Advances in neural implants, driven by progress in neuroscience, biomedical engineering, and artificial intelligence, have revolutionized fields like prosthetic control and the treatment of neurological disorders. Technologies such as deep brain stimulation (DBS) have demonstrated significant efficacy, while neuro prosthetics combined with artificial intelligence promise more sophisticated neural decoding systems. However, these developments raise critical ethical questions, particularly regarding their impact on freedom of thought and choice (García-Rairán & Clavijo-Montoya, 2021).

Neuro implants may create significant inequalities, not only in access to these technologies but also in the ability to resist their imposition. In contexts of social or economic pressure, individuals might be coerced into adopting neural devices, compromising their freedom of choice (Fukuyama, 2002). Moreover, the introduction of brain-machine interfaces raises questions about who controls neural data. If this information is hacked or used without consent, individuals could lose not only their privacy but also their autonomy over their own thoughts.

The ability of brain-machine interfaces to interpret neural signals and interact directly with the human brain raises concerns about the integrity of mental autonomy.

If such devices are used to influence or manipulate thoughts, individual freedom could be severely compromised. The use of technologies to modify or control brain functions risks turning individuals into designed artifacts, violating their authenticity and autonomy (Habermas, 2003).

The mass collection of neural data, essential for the functioning of many neuro technologies, also presents risks to freedom of thought. Intimate information about individuals' neural patterns could be used to influence behaviors or decisions, creating an environment conducive to manipulation and external control. Surveillance technologies that collect and analyze personal data may undermine privacy and self-determination (Zuboff, 2019).

Another crucial issue is the vulnerability of neural devices to invasions or misuse by third parties. While implants promise to enhance cognitive capacities, they also create the possibility of "brain hacking," where neural information could be accessed or altered without consent. This direct threat to mental sovereignty raises concerns about ensuring freedom of choice in an environment where thoughts can be externally influenced (Musk, 2020).

The use of these technologies in corporate or governmental settings may introduce new forms of coercion. Companies might require employees to adopt neural implants to boost productivity, thereby limiting individual autonomy. Potential interference with freedom of thought goes beyond direct control. Neural interfaces that offer suggestions or automatically complete actions could subtly shape users' thought patterns and decisions. This "automated assistance" might reduce critical reflection, making individuals increasingly dependent on technology for complex decision-making (O'Neil, 2016).

Given these challenges, it is essential that the development of neuro technologies be accompanied by rigorous regulations. Protecting freedom of thought and choice must be a priority in any ethical debate on the implementation of these technologies. The European Union, for example, with its General Data Protection Regulation (GDPR), has taken a step in the right direction by imposing restrictions on the collection and use of sensitive data (Zuboff, 2019).

Health monitoring devices, such as smartwatches, promise greater control over physical well-being but store sensitive data that could be used for surveillance or discrimination. An example is the use of COVID-19 tracking apps during the pandemic, which sparked debates about the fine line between public protection and invasion of privacy (Zuboff, 2019).

Enhancement technologies also create a paradox of choice: as enhancement options multiply, individuals may face social pressure to adopt technologies, even against their will. Fukuyama argues that freedom can be undermined when enhancement choices become normative, creating a society were refusing to enhance oneself leads to social or economic exclusion (Fukuyama, 2002). This pressure is evident in educational and professional contexts. The use of cognitive enhancement drugs is increasingly common among students and high-performance professionals. While these drugs promote performance freedom, they also create inequalities between those who have access to them and those who opt not to use them, perpetuating a cycle of artificial competition.

The pursuit of enhancement also impacts collective freedom. The introduction of autonomous weapons controlled by artificial intelligence illustrates how innovative technologies can be used for control and repression rather than liberation. The use of drones for mass surveillance by governments exemplifies how technologies developed to enhance security can restrict civil liberties (Zuboff, 2019). Additionally, the development of "smart cities" equipped with sensors and cameras to monitor human behavior raises concerns about a dystopian future where individual freedom is sacrificed for efficiency and order.

In light of these challenges, adopting an ethical approach to the use of enhancement technologies is crucial. Habermas suggests that the debate on transhumanism should be conducted within the framework of a social contract, where the limits of technological intervention are collectively defined (Habermas, 2003). This perspective seeks to preserve freedom while promoting advancements. Regulations are also necessary to protect privacy and autonomy. Organizations such as the European Union have already implemented guidelines like the General Data Protection Regulation (GDPR) to ensure that data-based technologies respect individual rights.

The paradox of enhancement and freedom reflects the inherent ambiguity of technological progress. While innovations promise to liberate humanity from physical and cognitive limitations, they also introduce new forms of dependency, control, and surveillance, threatening individual autonomy. This scenario leads to an equally critical question: how do these technological advancements impact the intrinsic value and dignity of human beings? Ultimately, enhancement not only challenges freedom but also calls into question what it means to be human in a world increasingly shaped by technology.

3. The paradox of human dignity: technological enhancement or dehumanization?

The quest for human enhancement, while promising in terms of evolution and progress, brings to light the paradox of how to preserve dignity while utilizing technologies that challenge the boundaries of nature. Dignity, understood as the intrinsic value of each individual, can be threatened by the instrumentalization of the body and mind, transforming humans into objects of enhancement and exploitation. This issue not only resonates with contemporary debates but also echoes philosophical and mythological traditions.

The United Nations emphasizes that human dignity is an inalienable value that must be preserved under all circumstances (UN, 2024). However, transhumanist technologies that aim to modify human essence may be seen as an affront to this principle. Genetic modifications to improve physical or cognitive traits raise questions about what it means to be human and the extent to which such interventions respect the ethical boundaries of science.

The myth of Pygmalion, as narrated by Ovid, reflects these dilemmas. The story of the sculptor who idealizes a perfect statue and sees its perfection compromised when it is brought to life by divine intervention symbolizes the risks of interfering with the natural aspects of humanity. This myth anticipates contemporary debates

about authenticity and the potential impacts of human enhancement on dignity (Ovid, 2008).

Conversely, proponents of transhumanism argue that advanced technologies can strengthen human dignity by overcoming physical and cognitive limitations. Bostrom (2003) asserts that human enhancement can eradicate conditions that undermine dignity, such as incapacitating diseases. Genetic editing, exemplified by CRISPR, offers opportunities to correct hereditary diseases and improve quality of life, enabling millions to live with greater autonomy (Doudna & Sternberg, 2017).

Bionic prosthetics and exoskeletons exemplify how enhancement can restore dignity by returning mobility and autonomy to individuals with disabilities (Herr, 2014). These technologies demonstrate that progress can indeed align with dignity by promoting greater freedom and control over one's life. However, these advancements also bring risks that cannot be ignored.

One such risk is the instrumentalization of the human being. Genetic interventions aimed at enhancement can compromise autonomy and dignity by shaping human characteristics according to external preferences. Dignity is closely tied to the unpredictability and authenticity of human development, qualities that could be lost in a scenario of deliberate and commodified choices (Habermas, 2003).

The practice of creating genetically modified babies is a clear example of instrumentalization. By selecting genetic traits to meet social or economic expectations, individuals are reduced to designed products, compromising their intrinsic value (Fukuyama, 2002). Moreover, inequality in access to these technologies could lead to a caste-like society where the enhanced enjoy privileges while the "unenhanced" face discrimination (Hughes, 2004).

Another significant challenge is the impact of transhumanist technologies on privacy. The mass collection of data by technological devices can transform individuals into sources of profit, dehumanizing them by reducing them to numbers in surveillance systems. Zuboff (2019) points out that this loss of privacy undermines dignity by subjecting individuals to constant monitoring.

Technologies such as facial recognition, while promoting security, have also been associated with discriminatory practices. These systems can reinforce racial or political stereotypes, violating the freedom and dignity of marginalized groups (O'Neil, 2016). In sports, the use of bionic prosthetics, such as running blades, generates debates about merit and equality, questioning what it means to compete fairly (Morgan, 2018).

In education, learning tools based on artificial intelligence promise to personalize students' experiences but may overlook their human particularities. The reduction of education to a mechanistic logic threatens dignity and individuality, devaluing the complexity of human learning.

Furthermore, the impact of technology on the human mind raises significant concerns. Neural implants, such as Neuralink, promise to treat diseases like Alzheimer's and paralysis but may also create dependency and pave the way for external manipulation. Musk (2020) warns of the risks that governments or corporations might use these technologies to control thoughts and actions, compromising freedom of choice.

The pursuit of biotechnological perfection reflects unchecked pride that disregards the value of human imperfections. Sandel (2007) argues that dignity lies in accepting natural limitations, which promote connection and humility. Technological interventions should be guided by a species-wide ethic that preserves human dignity and autonomy (Habermas, 2003).

Another worrying aspect is the dependency on technological devices. In addition to compromising privacy, these innovations may be vulnerable to cyberattacks, exposing individuals to physical and psychological risks. This reinforces the need for stringent regulations to protect users' rights.

Inclusive public policies are crucial to democratizing access to transhumanist technologies. Government subsidies and public research can ensure that progress benefits society as a whole, reducing inequalities. International initiatives, such as the Oviedo Convention, highlight the importance of establishing ethical standards for the responsible use of innovations.

Global collaboration is essential in this context. Regulations like the General Data Protection Regulation (GDPR) demonstrate that technological advancement can occur without compromising fundamental rights. However, a collective effort is needed to balance innovation with social responsibility. Ethical debates on human enhancement must involve scientists, policymakers, and society at large. Only through comprehensive discussions will it be possible to ensure that advancements uphold fundamental values such as dignity, freedom, and equality.

The paradox of human enhancement reflects the ethical dilemmas of modernity. Transhumanist technologies promise to expand capacities and overcome limitations, but they also pose risks of exclusion and manipulation. From the myths of Prometheus and Pygmalion to discussions on CRISPR and artificial intelligence, the central challenge is to balance progress and dignity. Therefore, technological advancement must be guided by solid ethical values that ensure inclusion and preserve human essence. Only in this way can progress become an ally of dignity, ensuring it benefits all humanity without compromising the principles that define us.

Conclusions

Transhumanism presents itself as a promising frontier for overcoming human limitations, proposing technological advancements capable of eradicating diseases, restoring lost functions, and enhancing physical and cognitive capacities. However, the pursuit of human enhancement raises profound ethical dilemmas and significant social challenges. Technologies such as genetic editing and neural implants offer potential benefits but also carry the risk of deepening social inequalities, transforming choices into obligations, and compromising fundamental values such as dignity and freedom.

One of the main challenges of transhumanism is accessibility. Innovative technological innovations, such as bionic prosthetics and brain-machine interfaces, are often prohibitively expensive, restricting access to the most privileged populations. This scenario creates deep divisions between an elite able to benefit from these innovations and the majority who remain excluded. As a result, technological progress,

rather than promoting equality, risks reinforcing and amplifying existing social and economic disparities.

Beyond economic concerns, transhumanism raises ethical questions about freedom of choice. In competitive markets such as employment, enhancement technologies may shift from being optional to becoming implicit requirements. Individuals could feel pressured to adopt these innovations to maintain relevance, transforming what should be a free choice into an obligation imposed by social or professional contexts.

Another critical issue is the impact of these technologies on individual autonomy. Neural implants and other cognitive devices have the potential to enhance cognition and treat neurological diseases, but they also open the door to subtle forms of external control. Governments or corporations could exploit these innovations to manipulate behavior, compromise privacy, and reduce the human mind to an object of technological oversight. Such a scenario threatens one of the pillars of human dignity: sovereignty over one's own thoughts and decisions.

Inequality in access to transhumanist technologies also raises the possibility of a society stratified by artificial capacities. Individuals who can afford these advancements will enjoy disproportionate advantages compared to those who cannot, consolidating a "biological caste" system. This stratification could exacerbate social divisions, undermine principles of equality, and make overcoming economic and cultural barriers even more difficult.

Privacy also becomes a central concern in this context. Many transhumanist devices rely on neural and personal data to function, fostering an environment conducive to the mass collection of intimate information. This data can be exploited for commercial or political purposes, exposing individuals to manipulation and cyberattacks. The lack of robust regulations to protect privacy could turn technological innovations into tools of surveillance and control.

In light of these challenges, it is essential that the development of transhumanist technologies be accompanied by ethical regulations and inclusive public policies. Government subsidies, public research, and international agreements can help democratize access to these innovations. Initiatives such as the Oviedo Convention and the General Data Protection Regulation (GDPR) demonstrate that it is possible to establish global standards that limit abuses and promote the responsible use of technologies.

Finally, technological progress must be guided by fundamental ethical values, such as dignity, freedom, and equality. The pursuit of human enhancement should prioritize inclusion and respect for diversity, avoiding the instrumentalization of the body and mind. A collective effort is indispensable to balance innovation with social responsibility, ensuring that transhumanism's advancements benefit all of humanity and uphold the principles that define our essence.

References

Besnier, J.-M. (2009). *Demain les posthumains:* Le futur a-t-il encore besoin de nous? Paris: Hachette Littératures.

Bostrom, N. (2003). Transhumanism and enhancement. *Journal of Evolution and Technology*, 14(1).

Bostrom, N. (2014). *Superintelligence*: Paths, dangers, strategies. Oxford: Oxford University Press.

Cave, S. (2020). The genome genie is out of the bottle: CRISPR and the future of editing humanity. *Nature Human Behaviour*, 4(6), 556–561.

Conselho da Europa. (1997). *Convenção para a proteção dos direitos humanos e da dignidade do ser humano face às aplicações da biologia e da medicina*: Convenção de Oviedo. Available at https://rm.coe.int/168007cf98. Accessed on Dec. 23, 2024.

Doudna, J. A., & Sternberg, S. H. (2017). *A crack in creation*: Gene editing and the unthinkable power to control evolution. Boston: Houghton Mifflin Harcourt.

Ésquilo. (2013). *Prometeu acorrentado* (M. da Gama Kury, Trad., 3ª ed.). Rio de Janeiro: Zahar.

Fukuyama, F. (2002). *Our posthuman future*: Consequences of the biotechnology revolution. New York: Farrar, Straus, and Giroux.

George, A. (2003). *The epic of Gilgamesh*. London: Penguin Books.

Gilbert, F. (2018). *Ethical challenges in deep brain stimulation*.

Habermas, J. (2003). *The future of human nature*. Cambridge: Polity Press.

Harari, Y. N. (2016). *Homo Deus*: A brief history of tomorrow. London: Harvill Secker.

Hughes, J. (2004). *Citizen cyborg*: Why democratic societies must respond to the redesigned human of the future. Boulder: Westview Press.

Hupffer, H. M., & Berwig, J. A. (2020). A tecnologia CRISPR-CAS9: da sua compreensão aos desafios éticos, jurídicos e de governança. Pensar: *Revista de Ciências Jurídicas*, 25(3).

Kass, L. R. (2002). *Life, liberty, and the defense of dignity*: The challenge for bioethics. San Francisco: Encounter Books.

Kemp, M. (2019). *Leonardo da Vinci*: The marvellous works of nature and man. Oxford: Oxford University Press.

Kurzweil, R. (2005). *The singularity is near*: When humans transcend biology. New York: Viking.

Morgan, W. J. (2018). The moral complexities of prosthetics in sports. J*ournal of the Philosophy of Sport*, 45(1).

Musk, E. (2020). *Neuralink and the brain's magical future*. Wait but why? Available at https://waitbutwhy.com/2020/07/neuralink.html. Accessed on Jan. 4, 2025.

O'Neil, C. (2016). *Weapons of math destruction*: How big data increases inequality and threatens democracy. New York: Crown.

OMS. (2021). *WHO issues new recommendations on human genome editing for the advancement of public health*. World Health Organization. Available at https://www.who.int. Accessed on Dec. 23, 2024.

OECD (2024) *Brazil advances in gender-sensitive budgeting with the launch of new guides and an OECD report*. Available at https://www.onumulheres.org.br/noticias/brasil-avanca-em-orcamento-sensivel-a-genero-com-lancamento-de-novos-guias-e-relatorio-da-ocde/. Accessed on Dec. 23, 2024.

Ovídio. (2008). *Metamorfoses* (P. F. Alberto, Trad.). Lisboa: Gulbenkian.

Platão. (2006). *A República* (M. F. M. Gomes, Trad., 3ª ed.). São Paulo: Martins Fontes.

Sandberg, A., & Bostrom, N. (2011). *Cognitive enhancement*: methods, ethics, regulatory challenges. Science and Engineering Ethics, 17(3), 311–341.

Sandel, M. J. (2007). *The case against perfection*: Ethics in the age of genetic engineering. Cambridge: Harvard University Press.

Shelley, M. (1818). *Frankenstein*. London: Lackington, Hughes, Harding, Mavor & Jones.

Sloterdijk, P. (2009). *Rules for the human zoo*: A response to the letter on humanism. Environment and Planning D: Society and Space, 27(1), 12–28.

UNESCO. (2005). *Universal Declaration on Bioethics and Human Rights*. Paris. Available at https://unesdoc.unesco.org. Accessed on Dec. 23, 2024.

United Nations. (1948). *Universal Declaration of Human Rights.* Available at https://brasil.un.org/sites/default/files/2020-09/por.pdf. Accessed on Dec. 28, 2024.

United Nations. (2023). *All human beings are born free and equal in dignity and rights*. Available at https://brasil.un.org/pt-br/255368-todos-os-seres-humanos-nascem-livres-e-iguais-em-dignidade-e-direitos. Accessed on Dec. 23, 2024.

Warwick, K. (2004). *I, cyborg. Champaign*: University of Illinois Press.

World Bank. (2021). *World Development Report 2021*: Data for better lives. Washington, DC: World Bank.

Zuboff, S. (2019). *The age of surveillance capitalism*: The fight for a human future at the new frontier of power. New York: PublicAffairs.

— 4 —

Neurotechnology, ethics and human rights: thinking about work, mental health inequities and eugenics

MONIQUE PYRRHO
Universidade de Brasília (Brazil)
pyrrho.monique@gmail.com

PEDRO ARAÚJO
Universidade de Brasília (Brazil)
pedro.jurid@gmail.com

Introduction

Neurotechnologies are blurring the boundaries between research and clinical practice, as well as between therapeutics and improvement. In a close exchange with clinical practice, neurotechnology research has been heavily dedicated to finding therapies for patients suffering from stroke and diseases such as Alzheimer's and Parkinson's (Cometa et al., 2022; Micera et al., 2020).

Neurotechnological developments can help guarantee human rights for people with disabilities and loss of sensory and motor functions. However, for this to happen it is necessary to think critically about what it means for technologies to ensure human rights. It is necessary to analyze technologies in their assumptions and purposes.

Neurotechnologies are convergence technologies. This means that they cannot be fully understood outside the context of interaction with other recent techno-scientific branches such as nanotechnology, biotechnology, information technology, and cognitive science (NBIC).

It is not only the boundaries between clinical practice and research that are being revised. Following a trend towards expanding health promotion as well-being, the distinction between therapy and enhancement is also increasingly challenged. In 2002, it was already announced that the NBIC convergence aimed to increase human performance and the productivity of nations. The fight against aging, the increase in

physical and cognitive capacities, human-machine and brain-to-brain interfaces were already facets of the promised revolution (Roco, Bainbridge, 2002).

The aim is to understand and intervene in brain functioning and/or Nervous System activities. The new technologies can be used for research or diagnostic purposes, or to interfere with and/or influence brain and/or nervous processes in therapeutic and improvement contexts (IEEE Brain, 2024).

A neural interface can be used to read and write information to and from the Nervous System. Electrodes or other sensors pick up electrical impulses that neurons send. Devices in direct dialogue with AI and data science decode those impulses and collect data, therefore reading minds. If they stop there, they are called open-loop devices. In addition to reading our neural activities, other kinds of neurotechnology can also send stimuli to affect brain and neural system functioning. In this sense they act in encoding, writing thoughts and behavior. These bidirectional devices are called closed-loop, and after decoding neural function they can automatically operate and alter brain function.

Devices and techniques are often classified into dichotomies described as open X closed loop, read X write, and monitoring X modulation of neural functions, and can be understood according to two possibilities of relating to nervous system functioning: collecting neural data and interfering in nervous processes (IEEE Brain, 20204).

Neurotechnologies which act by altering neural activities can be classified as: neuromodulation technologies, such as neural stimulation to reduce Parkinson's tremors or treat chronic pain; neuroprosthetics, which replace or restore motor or sensory functions, such as cochlear implants; and brain-machine interfaces, which would provide the ability to control devices by directly reading neural activities (IEEE Brain, 20204).

In addition to restoring functions or abilities, neurotechnologies aim to improve cognition and physical constitution, with applications that promise to revolutionize education, work, national defense, sports, leisure, and consumption (IEEE Brain, 20204).

Neurotechnologies can offer new paths for understanding and treating Nervous System disorders as a whole, but it cannot be ignored that these interventions are based on the interaction between devices and the neural tissue of individuals. Even when successful, this interference in neural tissue and its functions can result in changes in the personality and self-perception of individuals, as has already been reported in patients with Parkinson's Disease treated with Deep Brain Stimulation (Müller, Rotter, 2017).

Private companies are currently collecting indirect data on our neural activity through biometric data with the aim of analyzing reactions to stimuli related to advertisements, products or brands. Data provides information on emotional states, attention levels, preferences and mental conditions, and often in ways which are not perceptible to individuals. Obtaining this information enables companies to create more targeted and efficient strategies, improving the consumer experience and increasing assertiveness in decision-making processes (Farahany, 2023).

1. Ethics and neurotechnology

Discussing the ethics of neurotechnologies from the perspective of their risks is challenging. It is difficult to imagine a risk that has not been associated with these technologies. They include (but are not limited to) tissue damage due to implanting devices in neuronal tissue; risks to privacy due to increased invasiveness and pervasiveness of surveillance mechanisms; risks to the autonomy of subjects whose perceptions, feelings and behaviors are mediated by devices; risks of increased injustice, related to both access to technologies and to the distribution of their benefits and risks.

1.1. Neurotechnologies and work

Neurotechnologies are already a reality in the workplace, and certainly a disturbing one. The advantage initially touted was that of using vigilance monitoring mechanisms only in cases where the workers themselves could be at risk. But how can we not imagine that what will be monitored is not the workers themselves in their rhythms and efficiency? And that other data, unrelated to the most immediate interests of the work, are not additionally being obtained? The issue with the use of data is that we are often not even aware of the data we produce... Data can generate other data when combined and aggregated. How can we ensure that data collection is limited in time and space, but also in scope, strictly to that which is related to the work context?

The competitiveness for job openings, which constitutes coercion on individuals who need a means of subsistence, the opacity regarding the data obtained, but also regarding the destination of the information collected in the workplace, irrevocably harms the autonomy and concrete capacity of workers to consent (ICO, 2023).

Neurodiscrimination in the workplace is also emerging as a growing ethical concern as neuroscientific technologies become more accessible and used by organizations. The use of neural data or other biometric data to infer mental states can be used to justify decisions related to performance, leadership capacity or stress resistance, often without considering contextual factors, resulting in marginalizing and exclusionary discriminatory practices. This type of discrimination becomes even more problematic when such information is used to guide decisions regarding selection, or given that brain data can be manipulated or misinterpreted, reinforcing stigmas and stereotypes about groups of individuals (Dastin, 2018).

These are the concerns about obtaining data from workers. What if their behaviors can be programmed? There is increasing interest not only in collecting neurodata from employees, but also on stimulation and direct intervention in their neural functions, capabilities and perceptions in work environments. Will our imagination be able to come up with ways to prevent this from being harmful to the subjects? And to social life?

The purpose of neurotechnology is to modulate behaviors through surveillance, but also through direct interference in neural activity. Interference in the workplace is always announced with the dual objective of promoting quality of life while increasing performance in work activities. However, constant monitoring and an increased

expectation of greater productivity are the main causes of mental health problems in workers.

Psychological disorders in the workplace develop insidiously, causing progressive damage that is often overlooked. In addition to having difficulty proving that labor is the cause of their illness, workers end up having their work capacity and productivity affected, which often results in dismissal (Dejours, 2018).

Adopting neurotechnological devices as a selection criterion may also imply a violation of candidates' autonomy, since their use would not be optional, and would end up being a mandatory condition for recruitment. Employee data collection and storage may harm workers' future careers, since this information may be commercialized and/or accessed by multiple employers or for purposes other than well-being in the workplace. It is possible that neurodata collected in this way may generate new forms of discriminatory practices, pressuring workers to conform to tiring standards, damaging their dignity and quality of life at work.

Autonomy seems to be at stake in a way that is perhaps unprecedented. While it is true that people can be coerced into using these devices due to competition for jobs, the use of these devices may result in invading individuals' thoughts and emotions, thereby not only violating their right to privacy, but also resulting in intrusions into their thoughts that will most likely not be perceived as external (Muhl, Adorno, 2023).

Therefore, continuous monitoring imposes a burden of a similar nature and increased power to the psychological workload, in which the constant pressure for productivity and submission compromises workers' quality of life and dignity. This type of surveillance brings to light the need to rethink the ethical limits on the use of control technologies in the workplace, considering the impact of these practices on workers' mental health and autonomy. It is essential to guarantee clear and respectable parameters for using these technologies in order to balance the interests of companies with the rights of workers, ensuring that the workplace does not become a space of surveillance and domination.

Neurotechnologies seem to emerge as potential responses to contexts as described by Dejours (2018), in which the ever-increasing demand for productivity has generated enormous costs for the mental health of the world's population. In this scenario, would neurotechnology be the best of both worlds by offering the possibility of simultaneously increasing productivity and well-being? Or, on the contrary, is it just a disguise to extract even more value from minds that are already reaching their limits in face of the growing corporate appetite for efficiency and productivity? Whose interests does the neurotechnological solution respond to?

Even if biometric data is not obtained by devices implanted in invasive surgical techniques, it can already generate a feeling of violation of the right to privacy. For example, a supermarket chain in the United Kingdom implemented a surveillance system using armbands to monitor workers' productivity in its warehouses in Ireland and the United Kingdom in 2013. The devices recorded descriptive gestures of stock handling and breaks, such as those intended for using the bathroom. Some facilities announced by the employer, such as eliminating the need for manual records, did not avoid strong rejection by workers who denounced the measure as a form of oppressive surveillance. Continuous and pervasive surveillance, coupled with the constant

fear of punishment for possible dips in performance, were described as causing mental suffering by some employees (Rawlinson, 2013).

The situation becomes even more complex when one considers that neurodata is highly sensitive, as it comes from the electrical activity of the nervous system. This data provides information about internal mental activity, something that is intrinsically linked to the essence of the human personality. The lack of control over such information can therefore put privacy and individual rights at risk (Cornejo-Plaza et al., 2024).

Although we are subject to the possibility that value extraction and privacy violations are increasingly undetected, it is important not to limit the ethical discussion to our justified concern about privacy. The ethical acceptability of a technoscientific proposal involves the intentions of an intervention, its assumptions, the power dynamics involved, but also the final results of the phenomenon for justice.

1.2. Mental health and inequity

Happiness, productivity, long life promises versus the risk of loss of identity, autonomy and intimacy. Would there be something else to discuss regarding the interface between neurotechnology, ethics and human rights?

To go beyond wonder and fear, the first aspect to be considered from an ethical standpoint of an emerging technology is whether the potential benefits outweigh the risks. To do so, it is necessary to ask whether scientific and technological strategies are effectively based on concern for the specific distribution of the population's malaise, nervous system diseases and mental disorders. In other words, do neurotechnology promises respond to the reality of our health problems?

The answer: apparently not. The report presented to the UN General Assembly in July of last year highlights the inseparable relationship between extreme poverty and mental health. According to the data presented, the association between extreme poverty and job insecurity is the main cause of mental suffering in the world and triples the chances of mental disorders such as depression and anxiety. Women are the most affected: they are 3 times more likely to suffer from these disorders when compared to men. The report concludes that the obsession with economic growth, in addition to having led to an increase in inequality and income accumulation, has several circular effects: the highest prevalence of mental disorders is among poor people and people with mental disorders have a reduced chance of escaping poverty. In other words, the obsession with economic growth in countries reduces labor productivity. It is counterproductive. The recommendation is that Member States improve income distribution and safeguard workers' rights (UN, 2024).

In a pioneering study, Khan and fellows (2017) demonstrate that multifactorial discrimination (meaning discrimination related to the intersectionality between multiple social categorizations, such as race, gender, sexual orientation and class), acts as a determinant of mental health. The intersectionality between stigmatized social statuses acted as a predictor of an increased number of stressful events during life, chronic stress, high depression scores, substance abuse, suicidal ideation and suicide attempts. According to the authors, the greatest morbidity due to mental disorders is concentrated in non-white and poor LGBTQ+ populations; these same people are

those who have less access to protective factors for mental health, such as social protection networks and access to health services.

Public policies and clinical approaches need to incorporate intersectional perspectives to understand mental health inequities. It is not possible to develop effective strategies to promote well-being and health without focusing on the mechanisms that cause mental distress, disorder and illness.

1.3. Neurotechnology, eugenics and colonialism

The third aspect of ethical relevance is the increased risk that neurotechnology will bring eugenic consequences if directed towards human enhancement. Identifying the eugenic nature of supposed human improvements, which at first seems like a serious accusation, has nevertheless been loudly affirmed and defended. Following a trend to mitigate first the moral controversy regarding enhancement, and more recently regarding liberal eugenics itself, defenders of the ethics of human enhancement and/or improvement have emerged from the shadows.

The public defense of techno-scientific interventions for human improvement purposes inaugurates a new phase with the report "Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science". According to the document, the convergence between technologies and their dual uses, meaning therapy and improvement, is the path to follow for scientific and technological development policies (Roco, Bainbridge, 2002).

Nevertheless, there was widespread rejection of the proposal at the time. Among the documents that can be understood as a reaction to this perspective is "Beyond therapy: Biotechnology and the pursuit of happiness" (President's Council on Bioethics, 2003). The search for the best for children, or for better children, carries risks:

> We must consider the meaning of the ends themselves: better children, superior performance, ageless bodies, and happy souls. Would their attainment in fact improve or perfect our lives as human beings? Are they-always or ever-reasonable and attainable goals? (...) When asked what they wish for their children, most parents say: "We want them to be happy," or "We want them to live good lives" - in other words, to be better and to do better. The desire is a fitting one for any loving parent. The danger lies in misconceiving what "better children" really means, and thus coming to pursue this worthy goal in a misguided way, or with a false idea of what makes for a good or happy child (President's Council on Bioethics, 2003, p. 293).

In Europe, "Converging technologies: Shaping the future of European Societies" (Nordmann, 2004) is also a reaction to the proposal of technological convergence as a way to overcome the human condition. The ideas of enhancement and transhumanism are rejected in this document in the name of many science fields that together collaborate towards a knowledge society guided by social and humanistic values of European tradition.

Criticism of technological enhancement of humans seems to be losing traction, though. Two decades later, important organizations such as UNESCO (2024) and IEEE (2024) are now stating: neurotechnology isn´t just about treatment; increased performance and efficiency at work, increased learning performance, combating aging, neurally stimulated intelligence and happiness are the goals. These are the promising technical possibilities for promoting human flourishing through neurotechnology. The problem is that identifying the search for the most beautiful, most

intelligent and most efficient as ways to make humanity flourish and attributing these objectives to science poses a risk to human rights.

However, the idea is not to explore imagination and identify which rights are threatened. The guiding ethical concern that we intend to highlight here is the understanding that the proposal for human improvement engenders a split in our understanding of common and equally shared humanity. Scientifically sanctioning the discourse on what are the supposedly universal ideals of intelligence, beauty and biological functioning with the aim of improving society involves choosing which human characteristics are desirable. Humans are hierarchized in terms of their biological attributes.

Furthermore, thinking that the answer to social problems is a society that is organized and ranked by biological parameters is a central part of the historical plot of eugenics. It is the supposed scientific validation of these value hierarchies of human attributes that supported eugenic discursive and political practices aimed at eliminating inferior attributes in humans (and also humans considered inferior). Nothing new under the sun. What is surprising is the growing acceptability of the notion of eugenics, which in its "liberal" form is currently described as a softened and acceptable version of eliminating undesirable traits.

Proponents of liberal eugenics intend to describe it as an exercise of democratic freedom, because it is not perpetrated by the State, but rather by individuals who aim to improve themselves or their offspring. Liberal eugenics is thus sold as a manifestation of autonomy (Agar, 2019). The idea that this manifestation, because it is not coercive, would be democratic in itself is obviously quite problematic. Practices which reinforce prejudices and give a scientific veneer to racist beliefs that violate human rights cannot be considered democratic simply because individuals express a supposedly autonomous desire to perpetuate them.

The fact is that this eugenic trend is expanding rapidly. A survey in the US in 2008 showed that 80% of those interviewed were against genetic engineering of humans. At the same time, Richard Hayes, then executive director of the Center for Genetics and Society, denounced that genetic interventions aimed at human enhancement pointed to a neo-eugenic future in which inequities and prejudices would be exacerbated and that even parents opposed to gene editing would feel compelled to participate in a techno-eugenic race so as not to leave their children behind (Pray, 2008).

After more than a decade later, tolerance to eugenics is much higher. A study conducted in the US in 2023 showed that 57% would genetically select embryos for both medical and non-medical purposes. Of those interviewed, 41% would genetically edit their children (Meyer et al., 2023). Whether this trend is inevitable (as some claim) (Dias, 2025), is another question. Although eugenics has never completely disappeared, the movement, which covered the globe between the end of the 19th century and the first half of the 20th century, was at some point perceived in its horrendous reality, and acted to contain itself. The hope is that this limit, highlighted by World War II, will be perceived sooner rather than later in this century.

The problem is that eugenics is difficult and controversial to identify. First, it is often discursively equated to its authoritarian and Nazi manifestation. It is dismissed and relegated to a now-outdated moment in history, even intentionally by those who wish to promote it without receiving criticism. Additionally, each country

imprints specific aspects of its history on its eugenic narratives. In some ways, and in general terms, eugenics corresponds to the idea that there are superior biological attributes which should be promoted and inferior attributes that should be eliminated from population contingents. What this race is, what these attributes are, how the country understands itself in this dynamic and how it pursues such ideals is the topic under dispute.

Brazil enters the history of eugenics as a mixed-race country. Thus, some elements stand out from the start. First, the mixed-race identity replaces its real multi-ethnic composition. In addition to giving rise to the myth of racial democracy, this narrative makes the existence of indigenous and black people a restricted part of an origin myth and not a reality of the present. In the eugenic hierarchy, Brazilians are the people who do not want to coincide with themselves, they want to be another and pursue this through miscegenation. They are concretely multi-ethnic, but discursively they become mixed-race, and through migration and reproductive policies, they project a future altered by increasing miscegenation until they become completely whitened (Souza, 2016).

The clearly undemocratic goals of whitening the population did not end with the end of state eugenics in Brazil. A 2018 Wall Street Journal article points to an explosion in demand for American semen in Brazil, which multiplied more than 30 times in the 5 years between 2011 and 2016. Single wealthy women (26%), but also homosexual female couples (36%) and heterosexual couples (41%) import frozen semen from the US, and they mostly choose (95%) semen from white men, 52% of whom are blue-eyed (Pearson, 2018).

It is worth remembering that the first article of the Universal Declaration of Human Rights (UN, 1948), which began as a response to the terrors of eugenics, does not state that we have the right to equality. It more strongly states: we are equal. We have human rights because we are all considered equally human. To be even more repetitive, the humanity equally shared among us is the very foundation of Human Rights.

The UNESCO recommendation draft on the ethics of neurotechnology assumes that human enhancement, and the use of neurotechnologies in children to increase academic performance, are ethically acceptable, if certain conditions are observed. In other words, there are conditions in which the existence of humans who are better than others, or children who are better than others, is acceptable from an ethical and human rights point of view (UNESCO, 2024). But... what conditions would these be?

In short, that technologies be accessible and used only with consent. In other words, autonomy and justice. The benefit of enhancement seems so evident that it does not need to be weighed.

The issue is that the ethical defense of new technologies has been built on expectations which are known to be unattainable. It does not take much investigation to find out. It is already known that such conditions are nothing more than enunciating appeasing expectations. In fact, it is not expected that security and justice criteria will be observed. This tendency to enunciate conditions that will not be safeguarded has also been observed in relation to Big Data and artificial intelligence; more specifically, it is known that guaranteeing privacy and autonomy, understood as a conscious and free decision, are impossible from the outset due to the way in which these technologies are developed (Pyrrho et al., 2022).

There can be no coercion, but neurotechnologies can be used for work purposes. The use of neurotechnologies must be an autonomous decision, but it can occur in children of any age. What real space is there for autonomous choices in a world obsessed with economic growth? What is the possibility of a child, student or worker not feeling coerced into using these enhancing technologies in increasingly competitive environments? More than anything, neurotechnologies are convergence technologies between neuroscience and information sciences in which big techs invest millions, if not already billions, but the answer to it seems to be that "countries must make efforts to make technologies accessible" (Unesco, 2024). The fact is that the conditions for the ethics of these neurotechnologies all seem unachievable in the settings in which they are outlined. The same thing that happens with Artificial Intelligence, for example.

Neurotechnology-based enhancement involves defining rationality, mental health, functional minds, productive minds, and enhanced minds, as well as healthy bodies, normal bodies, and enhanced bodies. All of these are defined based on culturally defined expectations and desires, but which come to be taken as true because they are scientifically sanctioned. These perceptions are not neutral, much less universal. These conceptions have historically been used to rank people, making certain forms of suffering valid and becoming the object of scientific attention and investment, while other conditions are silenced, marginalized, oppressed, and penalized (Rivera-Segarra et al., 2022; Kiwan, 2022)

Convergent technologies, as we call those disruptive and yet combined big technoscientific endeavors (as exemplified by neurotechnology, genetics and AI), are presented as universalist paths to realize supposedly universal desires about the future of technology and humanity. Is what informs these projects universal? Are the dreams of humanity's future that these technologies outline in the form of improved bodies, or conversely disembodied minds, in fact universal utopias? What are the values and intentions that lie behind these enhancements that we are seeking? What informs the purposes of the technologies we promote?

The risk in mental health is especially great. Colonial dynamics categorize what is superior, desirable, civilized, rational, and on the other hand, what is inferior, archaic, and natural. Rationality is superior in this hierarchy. Consequently, the natural is available and at its service. This same classification guided, explained, and justified colonial practices, embodying the civilizing mission. On one side, the developed, conquering, rational, modern world, with its technologies, expanding rationality, taking science throughout the world, and on the other an environment and a human in a natural state, all to be occupied, landscaped, educated, and civilized.

This worldview is not something of the past, but is continually recycled in practices which gain scale with expanding convergent technologies. What does neurotechnology enhancement mean in this context? It means that some mental states and capacities are superior and desirable, others less so. There is great danger here...

There is another dimension related to colonialism, human rights and neurotechnology, still focusing on peripheral, traditional and original populations. These bodies are often exoticized, and were and are still seen as targets of scientific curiosity. It is notable to briefly recall the case of the Karitiana indigenous people who had blood samples and other biological materials collected in the 1990s, and with the ex-

panded idea of a global human genome collection, discovered that some biotechnology companies were selling these samples. All of this was denounced by Santos and Coimbra (1996), and the subject of a parliamentary commission of inquiry into the issue of biopiracy in 2004... But it came to nothing. The companies claim that they do not sell the material, they only pass on the costs. This seems to be enough for the widespread availability of this material without consent to continue with more than a little impunity to this day; it is still based on these materials and derived information that ancestries have become something to be measured by DNA percentages (Santos, Maio, 2008; Santos, 2002; Reardon, TallBear, 2012).

Neurotechnologies are divided into Reading/Writing interventions. In other words, it is about reading the mind, but also writing in it. What does it mean to read the minds of indigenous people, *quilombolas*, and riverside dwellers? Will we assume that they are diverse, and will we once again create a neurodata bank of human diversity, a showcase of brains like the Soviet eugenicists did at the beginning of the 20th century? Or, in a more refined manner, are we interested in obtaining their knowledge? Their mental health processes? Or even, will we read and write in their minds based on the parameters of what we consider pathological/normal/improved in our ways of life?

Conclusion

Neurotechnologies can be useful for people with disabilities, but we need to listen to them. Otherwise, the solutions may increase already historic inequities instead of helping to promote their rights. If people have reduced mobility, is our technological dream that there will no longer be bodies with reduced mobility? Or will we think that disability is something that exists in bodies in society? Can we dream of a society that considers the movement of diverse bodies? (Shew, 2023).

Should we imagine that we should all use technologies to become increasingly independent, functional, productive, and cognitively enhanced? Or will we be allowed to desire more inclusive and collective ways of life capable of understanding that economic independence is a patriarchal model of explaining the world, as Melinda Hall teaches us in "The Bioethics of Enhancement: Transhumanism, Disability, and Biopolitics" (Hall, 2016)? Do these perfectly independent bodies and minds that we stubbornly pursue really exist? Or, to give the appearance of independence, do many male, white, cis-heterosexual, Western bodies outsource and transfer all the burden and necessary work to other bodies? Aren't these bodies, the basis of all productive processes, the ones normally described as economically dependent and underdeveloped? But then, who depends on whom?

In hers "Imagination: A Manifesto", Benjamin (2024) reveals that these power dynamics that produce technology also structure what we dream about. The utopia for some of a technological world with improved humans replicates a very concrete dystopia for others. While we dream of superior bodies and minds because they are supposedly more independent, rational and productive, bodies and minds are then inferiorized according to these same valuations. People and their minds are hierarchized. This is not the future, it is the present and the past. But it can get worse... much worse.

Finally, what can be done so that these technologies can contribute to a better future? A good start is to recognize that people dream different dreams. And that the utopia of the oppressor is the nightmare of the oppressed... For technologies to reflect the dreams of a society that respects human rights and our diversity, it is necessary to ensure that this same diversity can dream and take the reins of technologies too, from the very beginning.

References

Agar, N. (2019). Why we should defend gene editing as eugenics. *Cambridge Quarterly Healthcare Ethics*, 28(1), 9-19.

Benjamin, R. (2024). *Imagination: a manifesto*. WW Norton & Company.

Cometa, A.; Falasconi, A.; Biasizzo, M.; Carpaneto, J.; Horn, A.; Mazzoni, A.; Micera, S. (2022). Clinical neuroscience and neurotechnology: an amazing symbiosis. *iScience*; 25(10), 105124.

Cornejo-Plaza, M.I.; Cippitani, R.; Pasquino, V. (2024). Chilean Supreme Court ruling on the protection of brain activity: neurorights, personal data protection, and neurodata. *Frontiers in Psychology*, 15, 1330439.

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool*. Reuters. Available at https://www.reuters.com/article/idUSKCN1MK0AG/

Dejours, C. (2018). *A loucura do trabalho: estudo de psicopatologia do trabalho*. 6 ed. São Paulo: Cortez – Oboré,

Dias, A.M. (2025). *Bebês geneticamente modificados são tendência inevitável*. Folha de São Paulo, 26 jan 2025. Available at https://www1.folha.uol.com.br/colunas/alvaro-machado-dias/2025/01/bebes-geneticamente-modificados-sao-tendencia-tecnologica-irrefreavel.shtml

Farahany, N.A. (2023). *The battle for your brain*: defending the right to think freely in the age of neurotechnology. St. Martin's Press.

Hall, M. (2016). *The Bioethics of Enhancement: Transhumanism, Disability, and Biopolitics*. Lanham: Lexington Books, 2016.

Information Commissioner's Office (ICO). (2023). *ICO tech futures: neurotechnology*. Available at https://ico.org.uk/about-theico/research-and-reports/ico-tech-futures-neurotechnology/ 2023).

IEEEbrain. (2024). *Brain topics*. Available at https://brain.ieee.org/resources/brain-topics/

Khan, M.; Ilcisin, M.; Saxton, K. (2017). Multifactorial discrimination as a fundamental cause of mental health inequities. *International Journal for Equity in Health* ,16(1), 43.

Kiwan, D. (2022). Dis/abled decolonial human and citizen futures. *Citizenship Studies,* 26(4–5), 530–538.

Meyer, M.N.; Tan, T.; Benjamin, D.J.; Laibson, D.; Turley, P. (2023). Public views on polygenic screening of embryos. *Science;* 379, 541-543.

Micera, S.; Caleo, M.; Chisari, C.; Hummel, F.C.; Pedrocchi, A. (2020). Advanced neurotechnologies for the restoration of motor function. *Neuron,* 105(4), 604 – 620.

Muhl, E.; Andorno, R. (2023). Neurosurveillance in the workplace: do employers have the right to monitor employees' minds? *Frontiers Human Dynamics*, 5, 1245619.

Müller, O.; Rotter, S. (2017). Neurotechnology: current developments and ethical issues. *Frontiers in systems neuroscience*, 11, 93.

Nordmann, A. (2004). *Converging Technologies–Shaping the Future of European Societies*. Interim Report of the Scenarios Group, High Level Expert Group, 3. Available at https://pure.iiasa.ac.at/id/eprint/12590/1/Converging%20Technologies.pdf

Pearson, S. (2018). *Demand for American sperm is skyrocketing in Brazil. Explosive growth spurred by more wealthy single women and lesbian couples turning to U.S. donors*. The Wall Street Journal, 22 mar 2018. Available at https://www.wsj.com/articles/in-mixed-race-brazil-sperm-imports-from-u-s-whites-are-booming-1521711000

Pray. L. (2008). Embryo screening and the ethics of human genetic engineering. *Nature Education*, 1(1), 207.

President's Council on Bioethics. (2003). *Beyond therapy: biotechnology and the pursuit of happiness.* Available at http://bioethics.gov/reports/beyondtherapy/index.html

Pyrrho, M.; Cambraia, L.; Vasconcelos, V.F. (2022). Privacy and health practices in the digital age. *American Journal of Bioethics*, 22 (7), 50-59.

Rawlinson, K. (2013). *Tesco accused of using electronic armbands to monitor its staff: Supermarket grades employees on efficiency and can reprimand them for breaks, says ex-worker.* The independent 13 Feb 2013. Available at https://www.independent.co.uk/news/business/news/tesco-accused-of-using-electronic-armbands-to-monitor-its-staff-8493952.html

Rivera-Segarra, E.; Mascayano, F.; Alnasser, L.; van der Vem, E.; Martínez-Alés, G.; Durand-Arias, S.; Moro, M.F.; Karam, E.; Hernández-Torres, R.; Alarcón, S.; Ramos-Pibernus, A.; Alvarado, R.; Susser, E. (2022). Global mental health research and practice: a decolonial approach. *The Lancet Psychiatry,* 9(7), 595–600.

Roco, M.; Bainbridge, W. (2002). *Converging technologies for improving human performance*: Nanotechnology, Biotechnology, Information Technology and Cognitive Science. National Science Foundation and Department of Commerce, Arlington/ Kluwer Academic Publishers, Boston, MA/ London: Springer.

Reardon. J.; TallBear, K. (2012). Your DNA is our history: genomics, anthropology, and the construction of whiteness as property. *Current Anthropology*, 53(S5). S233–S245.

Santos, R.V.; Coimbra Jr., C.E.A. (1996). Sangue, bioética e populações indígenas. *Parabólicas, Newsletter do Instituto Socioambiental*, 20(3), 7.

Santos, R.V. (2002). Indigenous peoples, postcolonial contexts and genomic research in the late 20th Century. a view from Amazonia (1960-2000). *Critique of Anthropology,* 22(1), 81-104.

Santos, R.V.; Maio, M.C. (2008). Genótipo e fenótipo: qual "retrato do Brasil"? Raça, biologia, identidades e política na era da genômica. In: Pinho, O.; Sansone, L. (org). *Raça: novas perspectivas antroplógicas*. 2. ed. rev. Salvador: Associação Brasileira de Antropologia: EDUFBA, p. 83-120.

Shew, A. (2023). *Against technoableism*: rethinking who needs improvement. WW Norton & Company.

Souza, V.S. (2016). A eugenia brasileira e suas conexões internacionais: uma análise a partir das controvérsias entre Renato Kehl e Edgard Roquette-Pinto, 1920-1930. *História, Ciências, Saúde – Manguinhos*, 23(1), 93-110.

United Nations (UN). (1948). *Universal Declaration of Human Rights*. General Assembly, Resolution 217A (III), A/RES/217(III) (December 10, 1948). Available at https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations (UN). (2024). Report of the Special Rapporteur on extreme poverty and human rights: Olivier De Schutter. The burnout economy: poverty and mental health. Available at https://documents.un.org/doc/undoc/gen/n24/210/39/pdf/n2421039.pdf

United Nations Educational, Scientific and Cultural Organization (Unesco). (2024). *First draft of the Recommendation on the Ethics of Neurotechnology*. Available at https://unesdoc.unesco.org/ark:/48223/pf0000391444?posInSet=8&queryId=N-EXPLORE-91e7956e-8457-4c7c-88a4-e234ff9f52c9

— II —

POLITICS AND ECONOMIC IMPACTS OF NEUROTECHNOLOGY

— 5 —

Neurolaw in non-democratic countries

TIAGO GAGLIANO PINTO ALBERTO
Pontifícia Universidade Católica do Paraná
tiagogagliano@hotmail.com

CECÍLIA SANTANA FIGUEIREDO PINTO ALBERTO
Pontifícia Universidade Católica do Paraná
ceciliasfigueiredo2507@gmail.com

Introduction

Neuro-law is a field that began to gain recognition within the legal community, largely during the 21st century, due to the rapid technological advancements that were previously unimaginable. It is evident that numerous individuals around the world benefit from the development of technologies aimed at understanding and enhancing the human mind, with the potential to even reverse chronic biological conditions, such as Parkinson's disease, through Deep Brain Stimulation (DBS). However, the issue addressed in this article concerns the delicate balance between technological evolution and freedom of thought, specifically, to what extent mental privacy and integrity are guaranteed both internally and externally as this new way of thinking takes hold. Formal regulation is currently almost non-existent in countries with a democratic ideal, exacerbating the situation even further in cases where democracy is not the primary system of government. Non-democratic regimes are the focus of this research, serving as a focal point to understand their distinction in relation to democracy, their peculiarities, and their relationship with freedom of thought.

1. The evolution of democracy: theoretical perspectives and definitions

To specifically understand what a non-democratic regime is, it is first necessary to comprehend all the facets that have accompanied the development of democracy

and its particularities since antiquity. It is important to emphasize that democracy does not have a fixed definition, evolving in accordance with the historical context and theoretical framework adopted.

In the past, Plato, in his works *The Republic* (Pangle, 1988) and *The Laws* (Adam *et al*., 1907), said that democracy was a corrupted government, only above tyranny, as the *demos* (greek word to common people of an ancient Greek state) itself was hierarchically divided and not democratically, and a good orator in the *ágora* could convince anyone. In modernity, this conception was used by Voltaire and Hume (Rasmussen, 2014) to justify their dominant ideology. For Rousseau, in *The Social Contract*, democracy was where the common good corresponded not to the sum of individual wills, but to each one's renunciation of their particular interests in the name of the common good (Rousseau, 2017). It is worth noting that Rousseau saw Plato as a theorist concerned with the education and freedom of the citizen.

In the twentieth century, John Rawls, in his book *A Theory of Justice*, talks about "deliberative democracy" as a model of justice that seeks to protect individual freedoms and conditions of social equality, reconciling liberal values with popular sovereignty, almost like a relationship between techniques of individualization and the totalization procedures of Pastoral Power (Rawls, 2017). In direct opposition to these ideas, Chantal Mouffe, in her article *Democracy, Citizenship, and the Question of Pluralism*, defends "radical democracy," also known as the "agonistic pluralism model," which ensures that the primary task of democracy is not to eliminate multiple passions or relegate them to the private sphere, but to guarantee rational consensus by mobilizing these passions for the development of the democratic design (Mouffe, 1992).

Currently, for Pietro Costa in his book *Few, Many, and All: Lessons from Democracy*, with the war in the 1930s, democracy was placed at the center of a "philosophy of war" adopted by the powers against Nazism, socialism, and fascism (Costa, 2012). However, over the years, and from the Universal Declaration of Human Rights in December 1948, the multiple fundamental rights and freedoms (known in the twentieth century as social) founded, under the ashes of totalitarianism in Italy, Germany, and France, the new "constitutional democracy." Political theorist Arend Lijphart concurs with the definition of democracy as "government by the people and for the people," yet he introduces a significant perspective concerning governance in the face of public disagreement. The conceivable solutions entail either rule by the majority (majoritarian democracy) or governance that strives for the broadest possible consensus (consensual democracy) when such disputes arise. The directly related difference is that the "majoritarian democracy" model is extremely exclusive, competitive, and combative, while "negotiated democracy" is characterized by inclusion, negotiation, and compromise (Lijphart, 2003).

Finally, Professor Marina Slhessarenko comprehends that Linz's democracy was inspired by Joseph Schumpeter's conception and Robert Dahl's polyarchies (Slhessarenko, 2021). For him, democracy is a system that develops the "free formulation of political preferences through the exercise of basic freedoms of association, information, and communication, with the aim of enabling fair competition among leaders to validate their claims to govern at regular intervals by non-violent means." (Linz, 2000). There can be no doubt in asserting that the aforementioned is a reflec-

tion and distillation of the most significant fundamental rights present within the framework of the Democratic Rule of Law. The republican principle encompasses not only popular sovereignty and citizenship, but also political pluralism and individual rights in the face of any internal or external threats. It affirms the contemporary understanding of the establishment of moral and principled objectives to be adhered to, ensuring the construction of a free and just society, while simultaneously providing robust protection in the realms of liberty, equality, and human dignity.

The essential attributes of fundamental rights define their existence both in the political and legal domains, serving as the cornerstone of the entire modern democratic edifice. Their progressive development throughout history has been geared towards countering instances where the human person was subjugated, thus safeguarding the inalienability and universal non-renunciability of such rights. However, one may assert without hesitation that discrimination and distinction continue to yield pernicious effects in the present day, relegating the universality of these guarantees to a secondary status. Freedoms that were once considered inherent, with their protective mechanisms sidelined, now open the door to the ineffective safeguarding of fundamental rights such as freedom of thought. The right to form, maintain, and express opinions without interference loses its essential character and begins to coexist with regimes where the democratic status is either weakened or, in certain cases, entirely absent.

2. Types of Non-Democratic Regimes: A Comparative Analysis

Having examined the characteristics of democratic systems, it is imperative to now focus on the defining features of non-democratic regimes. Among the most widely recognized traits of such systems are the concentration of power, the absence of free and fair elections, the repression of political opposition, the systematic violation of civil liberties and human rights, the pervasive control of the media, a lack of transparency and accountability, the utilization of security forces to suppress dissent, and the cultivation of a personality cult around the leader. These elements fundamentally differentiate non-democratic systems from democratic governance, where individual freedoms, public participation, and political pluralism are essential principles.

From a theoretical standpoint, Plato criticizes democracy, characterizing it as an ineffective system, merely a grand oratory exercise. However, he concedes that it permits deliberation on what is advantageous or detrimental to the public, irrespective of the variances among individuals. This viewpoint emphasizes the intrinsic complexities of democratic governance, wherein open dialogue and the negotiation of diverse perspectives are fundamental, yet it also underscores the challenges inherent in reconciling conflicting interests within such a system. The fact is that it would never accept any kind of repression of the opposition present in non-democratic regimes. Furthermore, John Rawls believes in defending individual freedoms and conditions of social equality within his “deliberative democracy”, not existing a possibility that this theorist would accept the lack of freedom of expression or free elections. The fundamental rights brought to light with constitutional democracy do not match in any way with the brute use of security forces or the lack of privacy with

obsolete media control. Therefore, everything that regardless of the divergences of the concept of democracy developed above would never be accepted, hence the importance of understanding how democracy differs from non-democracy.

The superficial characteristics of non-democratic movements are detailed, but the existence of different types of these regimes with their respective characteristics are likewise weighty. It is clear that there are many around the world that have formed or disappeared throughout history, some with similar characteristics and others completely different (Kaiser, 1997). Currently, it's possible to define 6 distinct types, being: totalitarian, authoritarian, fascist, sultanist, and dictatorial, considering also the existence of consolidated dictatorships in emergency situations. The relation of neurolaw mostly concerns totalitarian and authoritarian regimes, thereupon their specifications will be developed.

Authors like Adam Przeworski use "authoritarianism" and "dictatorship" synonymously (Przeworski, 2019), even though there are certain characteristics that differentiate them, such as:

a) Authoritarianism usually maintains a façade of controlling legality, while dictatorship tends to annul any constitutional precept;

b) Authoritarian regimes usually allow a small and specific popular participation, while dictatorships completely censor it;

c) Authoritarianism can have the concentration of power by a leader or a small group, while in a dictatorship the power is entirely exercised by a dictator or military junta.

On the other hand, authors like Brooker and Tóth defend exactly the lack of conceptual precision on the term authoritarianism, using, therefore, equivalent terms like "non-democratic," "arbitrary," or "illegitimate" (Brooker, 2014; Tóth 2017). Regarding their main characteristics, sociologist Juan Linz defines three central elements for the development of an authoritarian movement (Linz, 2000):

a) Maintenance of forms of divergence, even if authorized in its political expression channel.

b) Low popular mobilization or controlled mobilization of citizens, generated by the absence of an ideology responsible for popular identification and adherence.

c) Existence of a leader or small group exercising limiting power.

The birth, development, and repercussion of totalitarianism resemble different branches encompassing different parts of the phenomenal world. Eric Voegelin found it more convenient to use *gnosis* (a Greek word meaning wisdom) to represent intellectual movements in the modern and contemporary world of the twentieth century. In correspondence to Alfred Schutz (Eccel, 2016) a great philosopher of the mind, the author guarantees that religious experiences assume a great dimension in political life, that is, he understands totalitarian movements as political religions. In ancient Gnosticism, Christian and Jewish religious symbols were the basis for the relationship between God and man. In contrast, modern Gnosticism, summarized in Feuerbach's theses, God became just a projection of what men make of themselves: the reported "murder of God." In his work *The New Science of Politics*, the author comments on the literal and restricted analysis of the individual, abandoning any

trace of individuality and spirituality so embraced in Ancient and Medieval Ages. By way of example, he cites: Condorcet's progressive man, Comte's positivist, Marx's communist, and Nietzsche's Dionysian, proving that spiritual benefits have become institutional freedoms advancing towards different political regimes led by intellectuals (Eccel, 2016).

Contrary to Voegelin's ideals, Hannah Arendt, in her work *The Origins of Totalitarianism*, believes that totalitarianism cannot be studied based on any other system, always seeking uprooting from all kinds of tradition (Arendt, 2007). Norberto Bobbio in his book *Dictionary of Politics, Vol I* comments: "According to Hannah Arendt, Totalitarianism is a radically new form of domination because it does not limit itself to destroying man's political capacities, isolating him in relation to public life, as the old tyrannies and despotisms did, but tends to destroy the very groups and institutions that form the fabric of man's private relations, making him a stranger to the world and depriving him even of his own self." (Bobbio, 2007). Thus, its acceptance and permanence essentially reside in its ability to absorb and mobilize everything around it, as the inserted subjects come to allow and understand as necessary all kinds of crimes against humanity, convinced by various artifacts. Hence, the author defines the three main characteristics of this regime as:

a) Ideology full of dogmatisms with no room for any kind of criticism, ensuring the mass movement of the population in search of a specific end to be achieved.

b) Terror is applied to guarantee a state of generalized insecurity, with control of the opposition and destruction of the differentiation between the public and private.

c) Extensive propaganda to spread messages, increasing fear among opponents and ensuring forced participation in government-controlled organizations.

It is worth noting that each of these characteristics can be found individually in non-totalitarian regimes (Slhessarenko, 2021); however, their association effectively characterizes this type of system. Additionally, the first recognized totalitarian regime was that of Benito Mussolini in Italy in 1922.

3. Case studies of thought suppression: the struggle for freedom in authoritarian regimes

Regardless of the type of regime, its specific characteristics, or its historical development, the freedoms provided for in neurolaw and it's byproducts are not guaranteed. The most effective approach to examine this relation would be through the presentation and analysis of real cases where these provisions are nullified.

During Nazi Germany, there was a resistance group known as the "White Rose," founded by Hans Scholl, a professor at the University of Munich, with the help of his sister Sophie Scholl, and Christoph Probst and his students. Between June 1942 and February 1943, the group distributed anonymous pamphlets and graffitied public walls with anti-Nazi slogans, such as "Freedom" and "Down with Hitler," denouncing crimes, political repression, and the persecution of Jews. After being reported by a university janitor who claimed to be loyal to the regime, the Scholl siblings and Christoph Probst were tried by the infamous judge Roland Freisler of the People's

Court and sentenced to death for treason on February 22, 1943, being guillotined on the same day. Although the execution of the "White Rose" members was widely publicized by the Nazi regime as a warning to potential dissidents, the courage of the students and professors has become a lasting symbol of moral and intellectual resistance against tyranny, embraced to this day.

In the Dominican Republic, during Rafael Trujillo's regime, the Mirabal sisters (Patria, Minerva, and Maria Teresa) were among the most important political activists opposing Trujillo's brutal regime, participating in a clandestine organization known as the "14th of June Movement." Minerva and Maria Teresa Mirabal, in particular, were imprisoned and persecuted numerous times for their active participation in opposing Trujillo's brutal force against any expression of thought different from the regime. On November 25, 1960, Trujillo ordered the assassination of the sisters, along with their driver Rufino de la Cruz, who were brutally beaten to death, and their bodies thrown off a cliff to simulate a car accident. The death of the Mirabal sisters is widely seen as a turning point in the struggle against Trujillo's regime; Less than a year later, on May 30, 1961, Trujillo was assassinated in an ambush). Interestingly, in 1999, the United Nations General Assembly designated November 25 as the International Day for the Elimination of Violence against Women in honor of the Mirabal sisters.

In Egypt, during Hosni Mubarak's rule (Arab Spring), there were various mass popular protests demanding political, social, and economic reforms. In response to the protests, primarily organized through social media, the Egyptian government ordered a nationwide internet shutdown on January 28, 2011. This occurred after Vodafone and other telecommunications operators received government instructions to suspend internet and text messaging services to opposers. The company claimed that it was operating under a legal and national security obligation imposed by the Egyptian authorities at the time. However, it was criticized by the international community and human rights advocates for collaborating with an attempt to suppress freedom of expression and the free flow of information, raising ethical questions about the role of telecommunications companies in situations of government censorship and human rights violations.

In Saudi Arabia, blogger and human rights activist Raif Badawi, founder of the "Free Saudi Liberals" website, advocated and promoted debates on religion and politics, as well as supporting social and political reforms. In 2012, Badawi was arrested and accused of "insulting Islam" and "disobedience to his father." In 2013, he was imprisoned and sentenced to 7 years in prison and 600 lashes, later increased in 2014 to 10 years in prison, 1,000 lashes, and a substantial fine. Raif Badawi's imprisonment and conviction highlight the severe restrictions on freedom of expression in Saudi Arabia, where any criticism of the government or religious authorities leads to extreme punishments. The case underscored the lack of separation between religion and state in Saudi Arabia and how it affects citizens' lives.

In Hungary, the Central European University (CEU) was founded in 1991 by George Soros to promote open and democratic society through education and research, offering high-quality programs in social sciences, humanities, law, and public policy. However, during Viktor Orbán's government, the university came under attack as part of a broader campaign against philanthropist George Soros. In April

2017, the Hungarian Parliament passed a new higher education law ("Lex CEU") that imposed conditions making the university's continued operation in Hungary impossible. In 2018, due to the inability to meet all the criteria, the university moved most of its operations to Vienna, Austria. International academic organizations, foreign governments, and the European Union criticized the measure, arguing that the law violated academic freedom and democratic principles.

The Taliban, since returning to Afghanistan in 2011, have imposed various restrictions on women's rights, particularly regarding education. In March 2022, the Taliban announced that girls' secondary schools would remain closed until further notice and that women were banned from attending public and private universities. The basis for this decision, presented to the international community, was a strict and literal interpretation of Islamic law. The restrictions had demoralizing impacts on the lives of Afghan girls and women, facing a future with limited opportunities in various spheres.

Senegal cut off internet access in the country following protests against the conviction of opposition leader Ousmane Sonko in February 2024. The Ministry of Communications, Telecommunications, and Digital Economy issued a statement explaining that the measure was due to the dissemination of subversive hate messages in a context of public order disturbance, with telecommunications operators obliged to comply with all government-imposed requirements. The measure was quickly criticized by organizations such as Amnesty International and Reporters Without Borders, reaffirming that authorities must stop using the repressive machinery to silence any dissenting voices and ensure freedom of expression for those detained (Samira Daoud). The paradigm of freedom and privacy was completely disregarded.

4. The intersection of thought control, neurolaw and non-democratic regimes

It can be observed that the absence of defense for the freedom of thought and its corollaries is not confined to a specific moment in history, a particular continent, religion, theme, or group of individuals. On the contrary, the sole commonality observed across all practical instances is their divergence from democratic principles. Any notion of democracy was thoroughly disregarded upon the establishment of these totalitarian and authoritarian regimes. It is crucial to emphasize that many tend to perceive the atrocities committed by non-democratic leaders as being directed solely at natural persons; however, this is not the case.

Preventing the emergence of critical thinking and stifling intellectual development has become one of the primary objectives of these leaders, as evidenced in Nazi Germany, Hungary, and under the Taliban regime. Moreover, with the advancement of technology, these regimes were forced to adapt by blocking any form of intellectual dissemination through the internet, as seen in the cases of Egypt and Senegal. Additionally, it is worth noting that, in numerous instances, social media platforms serve as one of the most effective means for individuals to connect with the global community, as seen in Saudi Arabia. These platforms also provide leaders with an opportunity to implement their policies within the international sphere, utilizing what

has been termed "Neurovisual Law" (Brodovskaya and Dombrovskaya, 2020) to exercise a high degree of persuasive influence.

Neurolaw is the field of investigation that unifies knowledge from neuroscience with topics that may be relevant to law, or rather, which rights cannot be effectively guaranteed in such systems. In some cases, the population agrees with or even likes the repressive regime applied. According to Hannah Arendt (Arendt, 2007), the political capacity of human beings is abandoned, and hyper-individuality is embraced. The term "mass of people" comes from the fragmentation of social bonds, lack of belonging, which is then compensated for by the connection with the system. Thus, a new hierarchical construction is developed through repressive education with hostile impulses and the insertion of false memories, one of the main reasons for generalized conformity. This is the specific aspect in which neurolaw plays a role in studying non-democratic regimes from their inception to their respective end.

5. The erosion of freedoms in times of crisis: case studies from U.S. history

The foundation of democracy, as previously discussed, has always been the individual and social liberties of citizens, as well as the guarantee of consensus in political matters. However, this priority tends to shift when the state faces moments of intra or extraterritorial crisis. At this point, one of the most significant criticisms of the democratic model of protection throughout its historical development arises. A prime example to illustrate this is the development of conflicts in the United States of America over the centuries.

On April 30, 1789, George Washington, the first president of the United States, was democratically elected, followed by his re-election in 1792. Despite facing no constitutional restrictions and being the first in U.S. history, George Washington chose not to seek a third term, demonstrating the viability of a peaceful transition and the avoidance of power monopolization. Thus, on March 4, 1797, his vice president, John Adams, was elected and assumed office. It is no coincidence that the United States is still referred to as "the greatest democracy in the world." Even though it was one of the first countries to adopt modern democracy with its Constitution of 1789, its history of inclusion and stable institutions still falters in defending liberties during times of crisis.

The years from 1789 to 1801 were critical in American history, marked by uncertainty over the survival of the new Constitution and the bitter political clash between the Federalists in power and the Republicans, who were immensely sympathetic to the French Revolution. In response, the Federalists enacted the *Sedition Act* of 1798, which essentially prohibited the publication of any "false, scandalous, or malicious writing" against the government of the United States. This action drew significant criticism against John Adams' administration, as any matter opposing the Federalists was erroneously deemed an "act against the United States government." In this case, the state's defense was not even related to national sovereignty but to the defense and security of a political party. It was not until the case of *New York Times v. Sullivan* in 1964 that the Supreme Court finally declared the *Sedition Act* unconstitutional.

After the United States began actively participating in World War I in April 1917, strong opposition among American citizens intensified. President Woodrow Wilson then declared that "any form of disloyalty must disappear," and with the support of Congress, enacted the Espionage Act of 1917. This act made it a crime for any person to "cause or attempt to cause insubordination, disloyalty, or refusal of duty in the military forces of the United States or willfully obstruct the recruiting or enlistment service of the United States." It is important to note that it becomes nearly impossible to criticize any form of hostility without first highlighting military actions, thus becoming an intelligent way to once again obstruct any opposition to freedom of thought, discussion, and criticism of how the government was directing its efforts.

Decades later, the Supreme Court finally began to provide some security for freedom of expression, until the Cold War began. Restrictions to ensure national security were implemented, including widespread loyalty programs, attempts to outlaw the Communist Party, and extensive state and federal legislative investigations of suspected "subversives". Additionally, the *Smith Act* made it "unlawful for any person to knowingly or willfully advocate, abet, advise, or teach the duty, necessity, desirability, or propriety of overthrowing or destroying any government in the United States by force or violence." In 1951, the Court ruled that these restrictions did not violate the First Amendment in any way.

Authors Geoffrey R. Stone and David A. Strauss categorize this phenomenon into three rhetorical effects: pretext effect, chilling effect, and crisis effect. First, the "pretext effect" occurs when representatives create restrictions based on something that indirectly helps them silence any criticism that displeases them. Second, the "chilling effect" is based on the premise that individuals are fully aware they will face sanctions if they do not adhere to the established restrictions. Third, the "crisis effect" refers to the tendency of the population and the government to panic during times of crisis, developing a demonization of anything deemed dangerous, subversive, disloyal, or unpatriotic.

Today, there are international regulations that make it more difficult to implement such measures as those applied throughout American history, such as Article 19 of the UDHR and Articles 19 and 20 of the ICCPR, which pertain to Freedom of Expression. While these international protections cannot entirely prevent such governmental actions, they significantly reduce their occurrence.

Conclusion

In conclusion, it is understood that neurolaw as a field of knowledge does indeed encompass non-democratic regimes, trying to understand their peculiarities and similarities. However, its ultimate goal, which includes freedom of thought and its secondary effects (privacy, freedom of expression), is completely disregarded, since the central objective of any authoritarian or totalitarian leader is order as the center, regardless of how it is achieved. Furthermore, despite the emphasis that democracy places on both individual and collective liberty, it is evident that, particularly in times of crisis, this principle is often entirely disregarded in favor of safeguarding national sovereignty. Consequently, the teleological convergence between democratic and non-democratic states has emerged as a primary focus of contemporary neurolaw,

along with the various means, technological or otherwise, that have facilitated the attainment of this objective.

References

Adam, J., *et al.* (1907). *The republic of Plato*. Cambridge University Press.

Arendt, H. (2007). The origins of totalitarianism. In: Lawrence, B.B. & Karim, A. (eds.) *On Violence: A Reader*. Duke University Press, pp. 417-443.

Bobbio, N., Matteucci, N., & Pasquino, G. (2007). *Dicionário de política* (Vol. 1, 11th ed., p. 1248). Brasília: Editora UnB.

Brooker, P. (2014). *Non-democratic regimes*: Theory, government and politics. New York: St. Martin's Press.

Brodovskaya, E. V., & Dombrovskaya, A. Y. (2020). Dysfunctions of the political parties' internet content management. *2020 International Conference on Engineering Management of Communication and Technology (EMCTECH)*, Vienna, Austria, 1-4. https://doi.org/10.1109/EMCTECH49634.2020.9261535

Costa, P. (2012). Few, many, and all: Lessons from democracy. In *History of democracy*. Federal University of Paraná.

Eccel, D. (2016). Eric Voegelin e o gnosticismo: Da estreita relação entre religião, política e os regimes totalitários. *Numen*, 18(2).

Fraenkel, E. (2017). *The dual state*: A contribution to the theory of dictatorship. Oxford: Oxford University Press.

Kaiser, A. (1997). Types of democracy: From classical to new institutionalism. *Journal of Theoretical Politics*, 9(4), 419-444.

Lijphart, A. (2003). *Modelos de democracias: Desempenho e padrão de governo em 36 países* (pp. 23-26). Rio de Janeiro: Civilização Brasileira.

Linz, J. J. (2000). *Totalitarian and authoritarian regimes*. Colorado: Lynne Rienner.

Mouffe, C., & others. (1992). Democratic citizenship and the political community. In *Dimensions of radical democracy: Pluralism, citizenship, community* (Vol. 1).

Oxford Companion to Politics of the World. (2001). (2nd ed., pp. 57-59). New York: Oxford University Press.

Pangle, T. L. (Ed.). (1988). *The laws of Plato*. University of Chicago Press.

Przeworski, A. (2019). A conceptual history of political regimes: Democracy, dictatorship, and authoritarianism. In *New authoritarianism: Challenges to democracy in the 21st century* (pp. 17-36).

Rasmussen, D. C. (2014). *The pragmatic enlightenment*. Cambridge University Press.

Rawls, J. (2017). A theory of justice. In *Applied ethics* (pp. 21-29). Routledge.

Rousseau, J.-J. (2017). *The social contract and the first and second discourses*. Yale University Press.

Slhessarenko, M. (2021). Funções do direito em regimes não democráticos do século XX. *Leviathan. Cadernos de Pesquisa Política*, p. 11.

— 6 —

Abuse of economic power in the age of neurotechnology

CAMILA PINTARELLI
State Attorney of São Paulo
ckpintarelli@gmail.com

Contents: Introduction; 2. Cognitive liberty; 3. Mental privacy; 4. Abuse of economic power in the age of neurotechnology; Conclusion; References.

Introduction

Power is inherent to any political community and stems from the human capacity to act collectively, which, in turn, requires the consensus of many regarding a specific line of thought, externalized into a common course of action (Lafer, 2018).[1]

In fact, power sustains the existence of the public sphere, the potential space of the world of appearances among individuals who act and speak, as seen in the Greek expression *dynamis* or the German *match*, which highlight the term's potentiality, always a potential for power—a dynamic, not an immutable, measurable, or reliable entity like strength or vigor (Arendt, 2007, p. 212). The notion of power arises from the interplay of potency, possibility, and realization, and the force necessary for this to happen. In discussing power, we are reflecting on a dynamic that moves from concentrated possibility to full realization, driven by an element of energetic force (Ferraz Junior, 2009, p. 19).

As something dynamic, power is a process of constant renewal and rebirth, transforming it into a true means of communication. Therefore, the only indispensable material factor for the generation of power is human coexistence (Arendt, 2007, p. 213). Power does not rely on the concrete submission of its recipient nor, imme-

[1] Ainda a esse respeito: "Nenhum povo resiste ao seu próprio processo de geração e à competição com outros povos se não conseguir manter em funcionamento seu processo de autoinspiração. O que aqui se designa como inspiração autógena significa, em breves palavras, o contínuo de técnicas climáticas etnosféricas. Por meio de etnotécnicas que se estendem por gerações, dezenas, centenas de milhares, talvez milhões de indivíduos, afinam-se a espíritos comuns superiores e a ritmos, melodias, projetos, rituais e cheiros peculiares. Por força de tais jogos formais, que engendram uma sensibilidade comum e prática, os muitos homens reunidos encontram sem cessar, mesmo em circunstâncias adversas, as provas de que devem permanecer juntos; quando essas provas perdem sua força, os povos desencorajados se dissolvem no interior de culturas mais fortes ou se decompõem em bandos desordeiros e grupos remanescentes estéreis". (Sloterdijk, 2016, p. 55).

diately, on the effect achieved by its holder.[2] Instead, it is a process of transmitting selected actions between the holder and the recipient, constructed and existing based on the premise of the other's potential to act. Even if this action aligns with the selectivity exercised by the holder, the recipient retains the ability to think and externalize these thoughts through words, transforming them into action, albeit within a reduced horizon of appearances.

Thus, holding power in a communication relationship does not equate to subjugating the other; rather, it neutralizes the other, the recipient. The recipient continues to have various possibilities for action, but these possibilities are neutralized to such an extent that only one or a few options appear available, even though others have not disappeared (Santos, 1985, p. 153).

As power relationships – like any communication relationship - are shaped by social systems, they acquire more refined nuances depending on the social system and the value system in which they operate.

These initial observations are relevant to understanding the dynamics of power in the economic context, where its manifestation occurs through what is called economic power. This is one of the most significant components of the globalized economy, constitutionally protected in Brazil, and stems from the combined interpretation of the principles set forth in the constitutional framework of the Brazilian economic order, enshrined in Article 170 of the Federal Constitution.

Because it can neutralize actions, the exercise of economic power requires calibration in both practice and legal frameworks to ensure that economic agents - especially consumers - continue to exercise their freedom of initiative and will. This hermeneutic and operational adjustment of economic power occurs through protective elements for the agents involved in a given transaction, such as safeguarding free enterprise, protecting labor values, and ensuring freedom of competition. These elements simultaneously embody the characteristics of principles, rights, and duties, working to protect consumers, guarantee equal opportunities for all agents, and, more importantly, decentralize power.

Furthermore, historically, these elements have ensured - factually or legally - a degree of informational symmetry among the actors in the economic process. This allows, as previously mentioned, freedom of choice to remain unneutralized amid layers of decision-making possibilities. In other words, they are fundamental elements for containing abuse of economic power, understood as the exercise of this constitutionally recognized right in a manner that clearly exceeds the limits imposed by its economic or social purpose, good faith, or accepted customs.

[2] "[...] Estamos, pois, diante de uma situação em que o poder é meio para a transmissão de seleções de ações para outra seleção de ações (e não seleção de motivos de ações para ações), no qual ambos os comunicadores são sistemas aos quais se imputam seleções como suas ações. Assim, o submetido (sujeito "passivo" do poder) é alguém de quem se espera que escolha sua própria ação, donde a possibilidade de auto-determinação. Só que neste pressuposto é que são dirigidos contra ele elementos de poder, como ameaças, no sentido de regulá-lo nesta escolha por ele realizada. Do mesmo modo o detentor do poder se auto-determina. Com isso, na relação de ambos fica postulada a possibilidade de uma previsível e "localizada" divergência. A transmissão de complexidade reduzida ocorre quando a seleção da ação de um é co-determinada pela seleção da ação de outro. O sucesso de uma ordem de poder depende assim do aumento nas diferenciações de seletividades ainda relacionáveis. Ou seja, uma determinada ordem de poder que não consegue aumentar as diferenciações, que insiste em limitá-las a poucas possibilidades, (...), terá que ver diminuída sua cota de poder e aumentada a quota de violência, para sustentar – artificialmente – a situação, substituindo poder por coação". (Santos, 1985, p. 153).

Informational asymmetry is one of the most classic examples of abuse of economic power. It constitutes a market failure that occurs when one economic agent possesses a greater quantity or better quality of information than others, leading to inefficient, irregular, or even undesirable economic decisions. That is to say, in such circumstances, one economic agent, due to the information they hold, can neutralize the possibilities for action and decision-making of other involved actors.

To date, the study of economic power and the concept of abuse of economic power has been centered on a social reality that overlooked the existence of technologies capable of interacting directly with the human brain. As a result, legal and social constructions on the subject have not considered the need to protect the rights of economic agents at a stage prior to the external manifestation of human will.

Currently, however, we are witnessing a novel aspect of technological expansionism - techno-cerebral expansionism through neurotechnology. Neurotechnology is capable of surpassing the boundaries of the human psyche and, in an unprecedented way, unveiling the human mind, ushering in a new and perhaps irreversible stage in the revolution of social communication.

Between 2010 and 2020, private-sector investments in the development of neurotechnology increased twenty-two times, from $331 million to $7.3 billion, with global funding for neurotechnology companies totaling $33.2 billion in 2020. Moreover, projections indicate that the neurotechnology device market will grow exponentially, with economic values rising from $11.3 billion in 2021 to $24.2 billion in 2027 (UNESCO, 2023).

The intense development of neurotechnology, with advancements in connectome-based interfaces between the brain and machines, is genuinely expanding the boundaries of human action.

It should not be overlooked that this evolution brings with it hope and expectations regarding its positive impacts on humanity, offering countless possibilities for healing and improving human life. This was exemplified in 2014, during the FIFA World Cup in Brazil, when a paraplegic youth performed the opening kick-off with the aid of the BRA-Santos Dumont 1 exoskeleton - a robotic suit controlled by the patient's brain activity, captured through a sensor-equipped cap.

However, the improvement and establishment of these technologies, along with their increasing commercial availability, highlight the often-invisible or unpredictable impact they may have on life assets, whose legal interpretation has so far been inspired by the analog life rather than the neurotechnological life.

Considering these issues in conjunction with the economic considerations introduced earlier, UNESCO recently reported that consumers' emotional responses, associated with personal preferences, can now be tracked through neurotechnology devices such as neuroimaging tools. These tools are being used to predict market outcomes, create digital phenotypes, enhance video game tools, and advance "neuromarketing" – the use of neural data for advertising and promotional purposes (UNESCO, 2023).

It is also worth noting that a significant portion of today's personalized advertising operates through what is known as *real-time bidding*. This is an automated mechanism for auctioning advertising spaces on websites and other platforms, which,

within seconds, shares personal data with numerous companies. This enables the ad displayed in that specific space to be the one most aligned with the user's previously collected information. According to a recent report by the Irish Council for Civil Liberties, this industry generated over $115 billion as of May 2022. It exposes an individual's activity and location, on average, 747 times per day in the United States and 376 times per day in Europe (ICCL, 2022; O'Callaghan et al, 2024). Moreover, this is an industry that will likely begin to leverage neurotechnology, giving a new dimension to this form of advertising and opening unprecedented fields of impact on consumer decision-making and other economic agents.

This unprecedented context calls for a reevaluation of the concept of abuse of economic power within the neurotechnology landscape, especially given the scarcity of academic reflections on the legal and economic consequences that the dissemination of such technological tools may bring.

This will be the focus of our analysis, starting with the legal remodeling of the concepts of freedom and privacy in light of the unavoidable advances of neurotechnology. We will demonstrate the emergence of a new form of informational asymmetry, namely, the asymmetry caused by access to neural data. Finally, we will argue for the need to broaden the legal understanding of abuse of economic power today, such that it encompasses, even hermeneutically, the idea that such abuse may be characterized under circumstances that infringe cognitive freedom and mental privacy.

To achieve the objectives of our study, we will avoid delving into conceptual and metaphysical discussions about neurotechnology and its legal consequences. This is because we believe we are living in a historical moment where the impacts of neurotechnological devices on human rights are already understood by significant segments of our society. Our task, therefore, is to initiate the indispensable in-depth analysis of the myriad legal scenarios already being and yet to be affected by the consolidation and commercial use of neurotechnology, among which is the abuse of economic power.

2. Cognitive liberty

Freedom is the result of a human process that presupposes the social context in the taxonomy of elements shaping its exercise and realization in the world of appearances; when the context changes, so does the way freedom manifests itself. In this sense, and considering the historical movements that have colored the notion of freedom over the past centuries, we can currently perceive it as a configuration of complementary values, in which subjectivity mirrors the social sphere, and the social sphere finds in subjectivity the foundational root of its manifestations (Reale, 1976). This aligns with the perspective of Amartya Sen, for whom freedom and development are interdependent and complementary notions.

The requirements for the realization of freedom are, therefore, dynamic and - beyond being interdependent with the world of appearances - also depend on the behavior of others. The autonomy and agency of others can infringe on an individual's freedom (Sen, 1992), either by obstructing their possibilities for action or by interfering with their will. This interference can compel the individual to act according to

someone else's intentions or subvert their will through clandestine methods (Sætra; Mills, 2022).

Informational transparency plays a crucial role in shaping the exercise of freedom. A deficiency in the provision of information—whether through its absence, an excess that generates confusion, a failure to convey its idea correctly to the recipient, or the falsity of its content—deprives individuals of access to all alternative courses of action they could adopt in a given circumstance. This can lead to behavior that, even if not coerced physically or morally, does not reflect the purest exercise of freedom.

Today, with the rise and consolidation of neurotechnology, new layers of human susceptibility emerge regarding the exercise of freedom, bringing it into a new dimension: cognitive freedom. This is understood as the right to self-determination over our brains and mental experiences, encompassing the freedom to think, access one's own brain, and preserve the thoughts and memories it contains (Farahany, 2019).

Ienca and Andorno (2017) argue that cognitive freedom is multidimensional, presenting itself simultaneously as both negative and positive freedom. It is negative freedom in the sense of referring to choices and the exercise of cognitive control and mental integrity by the individual themselves, ensuring the absence of obstacles, prohibitions, or violations by public or private actors. Conversely, it is positive freedom in the sense of having the ability to act and take control of one's mental life.

Cognitive freedom, as positive freedom, is closely related to the sense of agency. Often described as the feeling of being in control of a specific event, the sense of agency arises in situations where an individual intends to produce a specific outcome, and their body moves through voluntary brain activity to achieve the intended result in the world of appearances. In other words, it is the experience that connects internal intentions with their external outcomes, positioning the individual as the protagonist of their actions and their influence on the surrounding reality, thus creating a sense of control over the given circumstances (Cornelio et al., 2022).

A fundamental characteristic of the sense of agency, which distinguishes us as humans (Farahany, 2012), is what Harry Frankfurt (1971) called "second-order desires". Beyond intending to produce an outcome and moving toward it, individuals may simultaneously harbor other desires or wish for circumstances to unfold differently. No other animal possesses this self-reflective capacity to form second-order thoughts, enabling us to hierarchize our neural activity processes to classify what is desirable and what is not (Taylor, 1985).

Although cognitive freedom, in its aspect of freedom of thought, is enshrined in Article 18 of the International Covenant on Civil and Political Rights (ONU, 1966) and Article 9 of the European Convention on Human Rights (European Council, 1951), its doctrinal and jurisprudential development remains rare. Only recently has it begun to receive greater attention and scientific accuracy—perhaps due to the persistent skepticism regarding the technological possibilities of interfering with, manipulating, or modifying human thought.

In jurisprudence, the two primary paradigms of cognitive freedom emanate from the U.S. Supreme Court.

In 1927, in a dissenting opinion in the landmark case *Olmstead v. United States* (277 U.S. 438) – a precursor to interpreting the Fourth Amendment to the U.S. Constitution – Justice Brandeis arguably made the first jurisprudential reference to what we now call the right to cognitive freedom. In his visionary and bold position, he foresaw, even then, that advances in psychiatry and related sciences might one day enable the exploration of unexpressed feelings, thoughts, and emotions.[3]

In 1968, in the case *Stanley v. Georgia* (394 U.S. 557), Justice Marshall (1968) asserted that the entire legacy of the U.S. Constitution would rebel against granting the Government the power to control individuals' minds and thoughts.

In more recent scientific literature, the renowned Harvard University Professor Laurence H. Tribe (1978)[4] makes explicit reference to the prohibition of domination through the control of mental processes in his analysis of the Stanley *v. Georgia* case (394 U.S. 557).

A significant contribution to this topic came from the groundbreaking study of Richard Glen Boire, Executive Director of the Center for Cognitive Liberty and Ethics in Davis, California, titled *On Cognitive Liberty*. Boire examined the historical background, feasibility, and implications of discussing cognitive freedom considering the amendments to the U.S. Constitution. He emphasized that the invisibility of mental processes makes identifying potential violations a particularly challenging endeavor, although such violations exist in a manner comparable to violations of an individual's physical freedom (Boire, 1999). These concerns were revitalized in an article by Wrye Sententia, who argued that the freedom to control our own consciousness is the foundation for any other human freedom. According to Sententia, discussing cognitive freedom equates to updating the guarantee of freedom of thought considering neurotechnological evolution (Sententia, 1968).

While our brain activity can support significant clinical advancements and other possibilities for social inclusion, it is increasingly subject to technological influences capable of accessing, inferring, manipulating, processing, and emulating our neural data.

Since the success of neurotechnological tools relies on the simultaneous use of artificial intelligence and machine learning processes, we cannot ignore the impacts of algorithmic interaction on informational dissemination and its resulting consequences.

---

[3] "[…] Advances in the psychic and related sciences may bring means of exploring unexpressed beliefs, thoughts and emotions. "That places the liberty of every man in the hands of every petty officer" was said by James Otis of much lesser intrusions than these. To Lord Camden, a far slighter intrusion seemed "subversive of all the comforts of society." 2 Can it be that the Constitution affords no protection against such invasions of individual security? (…) The protection guaranteed by the Amendments is much broader in scope. The makers of our Constitution undertook to secure conditions favorable to the pursuit of happiness. They recognized the significance of man's spiritual nature, of his feelings and of his intellect. They knew that only a part of the pain, pleasure and satisfactions of life are to be found in material things. They sought to protect Americans in their beliefs, their thoughts, their emotions and their sensations. They conferred, as against the Government, the right to be let alone-the most comprehensive of rights and the right most valued by civilized men. To protect that right, every unjustifiable intrusion by the Government upon the privacy of the individual, whatever the means employed, must be deemed a violation of the Fourth Amendment." (TAFT, 1927).

[4] "[…] In a society whose 'whole constitutional heritage rebels at the thought of giving government the power to control men's own minds,' the governing institutions, and especially the courts, must not only reject direct attempts to exercise forbidden domination over mental processes; they must strictly examine as well oblique intrusions likely to produce or designed to produce, the same result (Tribe, 1968, p. 889).

In this context, Sætra and Mills (2022) recall that the degrees of sensitivity and accessibility to freedom are often illustrated by the metaphor of doors available for us to open and enter. In the technological landscape, technology and algorithms continuously rearrange these "doors" we encounter in virtual life, intentionally manipulating information—and, consequently, the available options for choice-thus shaping individual behavior. This dynamic exacerbates informational asymmetry by processing an enormous amount of individual data, allowing platforms to construct precisely the "doors" they desire us to open.

By shaping our virtual behavior and segmenting our field of action, the combination of informational selectivity and artificial intelligence leads to the development of inadequate decision-making processes. These processes impair freedom by determining which possibilities and opportunities align with the minimum that each individual would like to achieve (Sen, 2000). The ability to act persists but only within the framework of choices pre-determined by the platform.

Returning to the question of neurotechnology, we see that the consequences for cognitive freedom are even more disruptive. This is because these consequences unfold in two levels of understanding, based on the notions of sensitivity and accessibility to opportunities for action and results. These understandings are fundamentally rooted in Kant's second imperative (Kant, 2007), *i.e.*, the purpose the neurotechnological tool assigns to the brain and the human subject: whether it is an end or merely a means to achieve other objectives, even subliminally.

The first level of understanding of cognitive freedom in the neurotechnological context pertains to the promising benefits that neurotechnology has already brought and will undoubtedly continue to bring to humanity. This clearly positions the human subject as an end in themselves.

In this conception, the freedom of individuals is not only protected but is also truly redefined. It takes on such a broad meaning that it transcends the impulse for cognitive freedom and reaches the actual realization of human dignity. Neurotechnology here effectively serves as a means to ensure accessibility to opportunities for actions and results, as determined by the individual themselves.

However, neurotechnology - particularly brain-computer interfaces (BCIs) - while enabling cognitive freedom and advancing human dignity, also holds the potential to manipulate, alter, emulate, and even "hack" individual actions (UNESCO, 2023).

This brings us to the second level of understanding of cognitive freedom in the neurotechnological context. Here, neurotechnology directly interferes with the spectrum of human action, positioning the individual not as an end in themselves but as a means to achieve other tasks and goals through compromised actions.

It is precisely within this level of understanding that we can observe the abuse of economic power.

The dissemination of neurotechnology use can, as we shall see, enable economic agents to develop new and previously unimaginable ways of manipulating, inducing, and stimulating economic actions in other actors within the economic chain. These actions would likely not occur if there were sufficient information about the

potential of a given neurotechnology available for commercial sale or greater clarification about its impact on decision-making processes.

In this regard, Yuste et al. point out that the integration of neurotechnology into human beings can provoke involuntary behaviors or behaviors unknown even to the individuals themselves. They further caution that if multiple people can control neurotechnological devices through thought or if several brains are connected for collaborative work, our understanding of who we are and what we are truly doing may become severely compromised (Yuste, 2017).

With respect to aspects related to the sense of agency, particularly the notion of control, Steinert et al. note that machine learning tools integrated into neurotechnology are capable, at a certain stage of learning, of inferring, predicting, and acting on the user's intention. This is done by analyzing previous mental commands, enabling the neurotechnological device to independently initiate new commands without prior conscious human action (Steinert, 2019).

Today, anyone can acquire peripheral devices equipped with electroencephalogram (EEG) technology. Despite their evident limitations in accessing the brain's imagistic content, these devices can record and store electrical brain activity, converting it into computational data.

In fact, the electroencephalogram (EEG) is a type of neurotechnology capable of capturing and recording raw brain data - slow waves (delta), medium waves (theta), and the three layers of high waves (alpha, beta, and gamma, the highest) - as well as the electrical signals from nearby muscles and ambient noise. These data are processed using artificial intelligence and filtered to extract relevant information, which is then returned to the individual (Farahany, 2023, p. 23) in the form of useful data-such as physical performance, sleep quality, emotional state, etc.-becoming increasingly common with the phenomenon of the "quantified self" (Farahany, 2023, p. 23).

Once captured and stored, raw data remain accessible and can be reanalyzed at any time to extract further information (Farahany, 2023, p. 23) and connect them to other non-neural data. By inference (Ienca; Malgieri, 2022), this enables even more insights into the individual.

It is worth noting that most personal-use peripheral devices equipped with neurotechnology currently available in the consumer market are, in some way, part of a broader technological ecosystem developed by the same manufacturer. These devices are invariably configured to electronically associate them precisely with a specific user. Consequently, the information processed by the neurotechnological device also feeds into and is influenced by other data, applications, records, and information within this broader technological set, ranging from phone data to health insurance records, bank accounts, and social media profiles.

This means that all our habits and routines can be monitored at the neural level to contribute to a comprehensive and unified record of information about us. Aside from privacy concerns, it is clear this unique circumstance ensures that the data provided by portable neurotechnological devices are almost never neutral. Instead, they are filtered through a myriad of information about our personality, which we continuously make available.

Thus, without formal acknowledgment, our behaviors begin to be conditioned by the logic of informational selectivity calibrated through algorithmic activity, which now incorporates data from the most critical element of human beings: brain activity.

The use of neurotechnology that positions the human subject as a mere host of computerized mechanical action directly impacts self-determination over our brains. This leads to consequences for sensitivity and accessibility to opportunities for action and behavior, resulting in the partial or total loss of freedom and agency.

Accessing brain activity via neurotechnology is thus provoking a profound reinterpretation of the concept of freedom. Freedom is no longer confined to exogenous factors external to the human subject but now encompasses, above all, mental processes. It becomes a genuine prerequisite for all other forms of human freedom, as well as other guarantees emerging in the techno-social context.

By perceiving freedom as a process of sensitivity and accessibility to alternatives for action and results - and as both a means and an end to achieve development with direct implications for the real opportunities afforded to people - cognitive freedom assumes critical importance. It is from cognitive freedom that all possibilities for human action emanate.

Without cognitive freedom, individuals enter life and economic relationships under conditions of inequality, manifesting a new form of abuse of economic power within a neurotechnological society.

3. Mental privacy

Beyond issues related to cognitive freedom, the manifestation of abuse of economic power in the neurotechnology era also invites us to reconsider the concept of privacy.

In his in-depth analysis of the phenomenon of privacy, Daniel J. Solove presents it as something culturally and historically contingent - that is, privacy varies according to time and culture (Solove, 2018). According to Solove, concerns about privacy arise from social intrusion, and thus its analysis should focus on the problems that create the desire for privacy, as well as the risks and harms such problems cause to society. Privacy concerns and protections do not exist in isolation (Solove, 2018). It is worth noting that the risk- and harm-based approach to privacy - at least in the context of artificial intelligence - is present in the European Union's *Artificial Intelligence Act*, approved in 2024 (European Parliament, 2019-2024). This act establishes three risk categories: limited risks, high risks, and unacceptable risks, with the latter being banned outright.

By focusing on the social problems that give rise to the need for privacy protection and the risks and harms they bring, we find that privacy should not be reduced to an individualistic understanding. Rather, privacy is a social and plural value capable of shaping our communities and producing profound effects on the structure of power and freedom (Solove, 2018).

Privacy thus promotes respect for individual personality and the ability to decide which aspects of one's life to share with others. Moreover, privacy aids in building

social trust, fostering control over one's life, and guaranteeing freedom of expression, which extends to ensuring broader social and political life (Solove, 2018). As Richard Neils eloquently states, privacy safeguards our identity and humanity, ensures our freedom and citizenship, and provides protections that allow us to trust others, thus enabling the construction of a better future (Richards, 2022).

One of the primary functions of modern technologies' design is communication: the transmission of messages and signals to those interacting with them. This function helps define our relationships and informs risk assessment and trust when dealing with others or with the technology itself (Hartzog, 2018). As a form of communication, technological design mirrors the structure of power, as previously discussed—facilitating the transfer of preselected actions to the end recipient. Design also shapes the digital economy by optimizing human tasks and calibrating the flow of information.

In this context, privacy—as an instrument to limit power—is naturally impacted by this dynamic. The informational flow, combined with the convenience technology offers in daily activities, affects individual perceptions of what privacy truly means (Muhl; Adorno 2023).

In a recent and masterful work analyzing privacy in the artificial intelligence environment, Solove highlights that *data scraping*—the extraction of data from one virtual location to another—is one of the most common practices in the technological space. He cites examples of companies extracting billions of photos from social media and other virtual profiles to feed large information repositories for facial recognition and photographic biometrics (Solove, 2024).

Solove argues that *data scraping* violates numerous privacy principles and standards. Individuals' data are appropriated by third parties without notification of the practice, consent or veto by the data owner, a clear and limited purpose, or data minimization. Consequently, there is no guarantee that the extracted data will be used exclusively for a specific purpose Solove, 2024). The extraction process is akin to issuing a "blank check," but without the knowledge of its owner.

This creates technological pathways for acquiring layers of individual informational data without consent - or even awareness - thus exposing new vulnerabilities related to human privacy.

Other forms of privacy violations may arise from this data-generation context, such as the creation of content that emulates human sensory capabilities. Examples include simulations of human behavior, the invention of harmful or false narratives about individuals or history, and so-called *deep fakes* - highly realistic content that exposes individuals to defamatory scenarios, often of a sexual nature. Privacy violations in these contexts occur due to the human subject's unawareness that they are interacting with an artificial tool, as well as through sensory manipulation derived from the quality of the content presented, which blurs the boundaries between reality and fiction (Solove, 2024).

These risks and harms assume new and previously unknown dimensions when we incorporate neurotechnology into this trajectory of technological development. This shifts the conversation beyond personal or sensitive data privacy to encompass mental privacy. This refers to the potential exposure of neural data – data derived

from the mental processes and activities of a human subject, relating to the brain's functioning or structure in an identifiable individual. This data contains unique information about a person's physiology, health, or mental states (OECD, 2022).

Ienca et al. define neural data as information concerning the structure, activity, and function of the human brain. This includes both direct measurements of brain structure, activity, or function and indirect measurements of their functional indicators. They further note that neural data are distinguished by their potential to be combined with non-neural elements – such as voice recordings, smartphone usage, or pre-existing psychological assessments – to form a set of information that enables inference and prediction through machine learning tools. Furthermore, they emphasize that neural data, while digitally storable, are not merely read-only; they are editable and can even be manipulated or rewritten via neurotechnology, particularly through electromagnetic stimulation and deep brain stimulation techniques (Ienca et al. 2022).

Neural data are thus directly linked to an individual's inner life, including their feelings, values, memories, identity, and personality. Unlike external information about human beings—though they may be externalizable—neural data are inseparable from the source that generates them: the individual's own neural activity (Ienca; Adorno, 2017).

Moreover, individuals generate a significant portion of neural data unconsciously or without any mental command prompting their generation (Steinert, 2019). This means that neurotechnology may access information from the subconscious and share data that would otherwise remain safeguarded within the mental sphere. This can occur both directly - for instance, through breaches of confidentiality - or by simply inferring such data via machine learning tools (UNESCO, 2023).

Notably, the Supreme Court of Chile issued a ruling in August 2023 that serves as the world's first judicial precedent analyzing privacy from the perspective of neural data. The court emphasized that technological advancements demand new protections for this dimension, which was previously confined to strictly medical contexts but is now accessible—and already being accessed—by individuals (Tercera Sala da Corte Suprema de Chile, 2023).

In this context, the current concept of privacy and its correlation with the concept of personal data, as enshrined in legislation worldwide, proves inadequate for addressing the challenges posed by neurotechnology. More than that, it falls short of the axiological, technical, and cultural contingencies presented by neurotechnological tools.

The ethical dilemma in these unprecedented circumstances of probing human essence revolves around striking a balance between the positive uses that access to brain privacy can provide for medicine and therapeutic clinics, and its use for less noble purposes (Farahany, 2022).

The impacts on mental privacy caused by neurotechnology, as with cognitive freedom, can be understood on two levels, based on the notion of social problems that give rise to the need for protection. These levels are fundamentally rooted in a comparison between the principles of solidarity and proportionality.

Adopting the doctrine of Robert Alexy (2012) and the concept of solidarity proposed by Jürgen Habermas (2013) – rooted in the construction of the nation-state and cooperative efforts based on a shared political perspective that transcends individual and local needs (Van Cleynenbreugel, 2018) –, we understand that the principle of solidarity can be integrated into the proportionality stage in its strict sense. This principle can then be weighed against mental privacy to address the two levels of understanding of mental privacy in the neurotechnological context.

As with cognitive freedom, the first level of understanding of mental privacy in the neurotechnological landscape pertains to values that go beyond individuality. In this case, it aligns with the notion of solidarity just discussed, as the partial relinquishment of mental privacy occurs to establish boundaries and realize the present and future of a collectively shared life context.

In other words, if we balance mental privacy with solidarity and find that the partial relinquishment of neural data privacy not only benefits the intersubjectivity shared by the individual but also benefits the individual themselves - making them both the subject and recipient of the benefits provided by neurotechnology's progress for their dignity and cognitive freedom—we fulfill another vision of Kant's second imperative (Kant, 2007). In such circumstances, the individual is preserved as an end in themselves, with neurotechnology serving merely as an instrument to safeguard human dignity.

However, the second level of understanding of mental privacy in the neurotechnological context involves scenarios where the balance between solidarity and mental privacy reveals a disproportionate outcome. It is precisely in this context that we can identify the abuse of economic power: neurotechnological tools enable economic agents to engage in a new form of surveillance—namely, the surveillance of neural information.

The term "surveillance," whose original meaning was closely associated with "watching over" or literally "keeping watch" over someone, now primarily involves the visibility of data (Lyon, 2022). The data we produce make us visible and facilitates learning, intentionality, routine, and control and influence purposes that underline surveillance by both government agencies and private actors.

Richards argues that humanity is living in the "age of surveillance," characterized by the creation of detailed records of our lives by the very technologies that have revolutionized how we live over the past decades (Richards, 2013).

If we live in the age of surveillance, and if this no longer depends on merely watching someone but instead relies on the visibility of the data we generate, neurotechnology undoubtedly introduces new and significant avenues for data collection, exposing mental privacy to unprecedented risks.

In this regard, there are already references to what is called the "age of brain surveillance" (Farahany, 2023). It is worth noting that the previously mentioned phenomenon of data scraping takes on even more concerning dimensions when applied to mental privacy.

It is not idle to recall the data we presented at the beginning of this study, according to which, as UNESCO (2023) reports, consumers' emotional responses related to personal preferences can already be tracked by neurotechnology devices, such

as neuroimaging tools. These tools are being used to predict market outcomes, create digital phenotypes, develop video game tools, and advance neuromarketing, which involves using neural data for advertising and marketing purposes.

Thus, neurotechnology opens the door for economic agents to begin surveilling neural information, enabling them to learn intentionally and systematically about individuals to influence and control their decisions (Lyon, 2007).

We believe that refining the second understanding of mental privacy will inevitably be accompanied—at least within the context of Brazilian legislation and regarding neurotechnological activities by private actors—by the concept of abuse of economic power, as outlined in Article 173, §4, of the Brazilian Constitution (Brazil, 1988) In these circumstances, economic activity would have exceeded the boundaries of its originally intended purposes, operating at levels of informational asymmetry capable of indicating domination not only of markets but also of the economic consumer. In this respect, the violation of mental privacy also infringes upon Kant's second imperative, as human beings, in such circumstances, are reduced to mere means for achieving the exclusive goals of economic agents.

4. Abuse of economic power in the age of neurotechnology

The advancement of neurotechnology has brought significant benefits, such as innovative medical solutions and improvements in human well-being. However, the growing commercial use of these technologies raises substantial concerns regarding the concentration of economic power and the abuse of this privileged position.

As we have observed, both freedom and privacy take on completely unprecedented legal perspectives in light of the commercial consolidation of neurotechnology, with social problems that go beyond the scope of this study. For us, what is important to highlight is the impact of this new techno-social trajectory on the economic order, understanding that the concept of abuse of economic power needs, at a minimum, to undergo hermeneutic reinterpretation.

The decision-making capacity of the human subject is no longer limited to exogenous influences that can strip them of their freedom of action by relying on informational flow. Today, we are faced with the real possibility of a true asymmetry of neuronal information, manifested both by the novel surveillance capabilities granted by neurotechnology to economic agents and by the direct manipulation of consumer preferences and decisions, as well as those of other economic actors. In the name of economic efficiency, this leads to a new form of abuse of corporate dominance.

The fusion of neurotechnology with artificial intelligence allows algorithms to determine human choices based on neural inferences. These decisions, cloaked in apparent innocence, can – and certainly will – be guided to favor commercial interests, directly interfering with the sense of agency, as previously discussed. The neutralization of options, typical of the abuse of economic power, takes on a new form: it is the cognitive process itself that becomes the object of manipulation.

Moreover, the concept of economic surveillance becomes particularly relevant in this unique context, as tools like brain-computer interfaces not only capture data for immediate use but also create permanent databases that enable continuous tracking. This places the human subject in a constant spiral of inference, anticipation, and

modeling of behaviors and preferences, completely undermining the fundamental principles of a fair economic order: fair competition and freedom of choice.

Abuse of economic power in the era of neurotechnology is already a reality, and we find it urgent that its legal framework, even through exegetical mechanisms, be improved to consider not only freedom of initiative and competition but also, and most importantly, the cognitive freedom and mental privacy of economic agents.

Conclusion

The substantial changes in the factual context brought about by neurotechnology challenge the normative intentionality previously expressed in response to social circumstances surrounding economic power and point to an apparent insufficiency of this very normativity.

We do not overlook the fact that the global society - particularly the Brazilian community - still lacks the concreteness and effectiveness of numerous fundamental rights already recognized in international and national instruments. The absence of public policies capable of implementing these rights makes the situation particularly concerning, especially when considering the idea of freedom as development, which we have discussed throughout this analysis.

However, we believe that this argument cannot be used as a reason to avoid addressing the regulatory need for a topic that is becoming increasingly tangible and sensitive in our reality. Its contours may only become visible when it is no longer possible to debate its ethical and normative consequences concerning the protection of the human person and, consequently, the economic agent.

Cognitive freedom and mental privacy emerge as essential pillars to ensure that scientific advancement does not result in a new form of oppression. In this context, the manifestation of economic power must be guided by these principles to prevent a new form of abuse of economic power, characterized not by external circumstances but by the improper use of the human being's neural activities themselves.

References

Alexy, Robert (2012). *Teoria dos Direitos Fundamentais*, 2o ed. São Paulo: Malheiros.

Arendt, Hannah (2007). *A condição humana*, 10º ed. Rio de Janeiro: Forense Universitária.

Boire, Richard Glen (1999). On Cognitive Liberty. *Journal of Cognitive Liberties* 1, v. 1, n. 1

Brasil (1988). *Constituição da República Federativa do Brasil*, 5 de outubro de 1988. https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Cornelio, Patricia. et al. The sense of agency in emerging technologies for human–computer integration: A review. *Frontiers in Neuroscience*, v. 16 https://doi.org/10.3389/fnins.2022.949138.

Corte Suprema de Chile. (2023), Rol N° 105.065-2023, 9 de agosto de 2023. Guido Girardi Lavín vs. Emotiv Inc. https://www.csjn.gov.ar/dbre/verNoticia.do?idNoticia=7595

Council of Europe. *European Convention on Human Rights*. Rome, November 4th, 1950. https://www.echr.coe.int/documents/d/echr/convention_ENG,

European Parliament. *Artificial Intelligence Act.* (2019-2024) P9_TA(2024)0138, https://artificialintelligenceact.eu/wp-content/uploads/2024/04/TA-9-2024-0138_EN.pdf

Farahany, Nita A. (2019). The Costs of Changing our Minds. *Emory Law Journal*. v 69, p. 75-110. https://scholarlycommons.law.emory.edu/cgi/viewcontent.cgi?article=1005&context=elj

Farahany, Nita A (2012). A Neurological Foundation for Freedom. *Stanford Technology Law Review*, v. p. 4–15. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=5322&context=faculty_scholarship

Farahany, Nita A (2023). *The Battle for Your Brain*: Defending the Right to Think Freely in the Age of Neurotechnology. New York: St. Martins Press - Kindle Edition.

Farahany, Nita A. (2023). Neurotech at Work. *Harvard Business Review*. v. 101, n 2, p. 43–48. https://hbr.org/2023/03/neurotech-at-work

Frankfurt, Harry G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, v. 68, n. 1, p. 5–20. https://philpapers.org/rec/FRAFOT

Ferraz Junior, Tércio Sampaio. (2009). Poder Econômico e Gestão Orgânica., em *Poder Econômico: direito, pobreza, violência, corrupção*, org. Tércio Sampaio FERRAZ JUNIOR, Calixto SALOMÃO FILHO, e Fábio NUSDEO, Barueri: Manole, 2009.

Habermas, Jürgen. (1989). Justice and Solidarity: on the discussion concernig 'Stage 6. *Philosophical Forum* 21. n 1, p. 32–52. https://philpapers.org/rec/HABJAS

Habermas, Jürgen (2013). Democracy, Solidarity and the European Crisis. In: GOZELIER, Anne-Marie et, al. (ed). Social European Report. p. 4-13. http://www.lasaire.net/upload/file/social%20europe%20journal/eBook%20def.pdf

Hartzog, Woodrow (2018). *Privacy's blueprint*: the battle to control the design of new technologies. Cambridge: Harvard University Press.

Ienca, Marcello; Adorno, Roberto (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy. v.* 13, n. 1, p. 5-32 https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-017-0050-1

Ienca, Marcello et al., Towards a Governance Framework for Brain Data. *Neuroethics*. v. 15, n 2, p. 15-20, https://infoscience.epfl.ch/server/api/core/bitstreams/d9c336e1-a2ff-4ce0-9f8d-4a3ae5dfb7e9/content

Ienca, Marcello; Malgieri, Gianclaudio. (2022). Mental Data Protection and the GDPR. *Journal of Law and the Biosciences* 9, n. 1, p. 1–19. https://academic.oup.com/jlb/article/9/1/lsac006/6564354

ICCL. Irish Council for Civil Liberties (2022). *The Biggest Data Breach*. https://www.iccl.ie/digital-data/iccl-report-on-the-scale-of-real-time-bidding-data-broadcasts-in-the-u-s-and-europe/

Kant, Immanuel. (2007). *Fundamentação da Metafísica dos Costumes*. Lisboa: Edições 70.

Lafer, Celso (2018). *Hannah Arendt:* pensamento, persuasão e poder, 3ª ed. Rio de Janeiro: Paz e Terra - Kindle Edition.

Lyon, David. (2022). *Pandemic Surveillance*. Cambridge, UK: Polity Press.

Lyon, David (2007). *Surveillance Studies: an overview*. Cambridge, UK: Polity Press.

Marshall, Thurgood (1968). Supreme Court of The United States. U.S. Reports: Stanley v. Georgia, 394 U.S. 557. *Retrieved from the Library of Congress - www.loc.gov/item/usrep394557*.

Muhl, Ekaterina; Adorno, Roberto (2023). Neurosurveillance in the workplace: do employers have the right to monitor employees' minds? *Frontiers in Human Dynamics* 5. https://doi.org/10.3389/fhumd.2023.1245619.

OECD (2019). *Recommendation of the Council on Responsible Innovation in Neurotechnology*, OECD/LEGAL/0457. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0457

O'Callaghan, Patrick. et al. (2024). The right to freedom of thought: an interdisciplinary analysis of the UN special rapporteur's report on freedom of thought. *The International Journal of Human* Rights. v. 28, n. 1, p. 1–23, https://doi.org/10.1080/13642987.2023.2227100.

Reale, Miguel. (1976). Considerações gerais sobre o projeto de Código Civil - Projeto de Lei no 634, de 1975", *Revista da Faculdade de Direito - Universidade de São Paulo*. v. 71, p. 25–58.

Richards, Neil (2022). *Why Privacy Matters.* Cambridge. Oxford University Press https://global.oup.com/academic/product/why-privacy-matters-9780190939045?cc=pe&lang=en&

Richards, Neil. (2013). The Dangers of Surveillance. *Harvard Law Review*. v. 126, n. 7.

Santos, Maria Celeste Cordeiro Leite dos (1985). *Poder Jurídico e Violência Simbólica*. São Paulo: Cultural Paulista.

Aætra, Henrik Skaug; MILLS, Stuart. (2022). Psychological interference, liberty and technology, *Technology in Society*. v. 69. https://www.sciencedirect.com/science/article/pii/S0160791X22001142

Sen, Amartya. (1992). Minimal Liberty. *Economica* v. 59, p.139–59. https://www.jstor.org/stable/2554743

Sententia, Wrye. (2004). Neuroethical Considerations: Cognitive Liberty and Converging Technologies for Improving Human Cognition. *Annals of the New York Academy of Sciences* 1013, n. 1 p. 221–228. https://pubmed.ncbi.nlm.nih.gov/15194617/

Sloterdijk, Peter (2016). *Esferas I: bolhas*, São Paulo: Estação Liberdade.

Solove, Daniel J. (2008). *Understanding Privacy*.Cambridge: Harvard University Press - Kindle Edition.

Solove, Daniel J. (2023). Data Is What Data Does: Regulating Use, Harm, and Risk Instead of Sensitive Data. *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4322198

Solove, Daniel J. (2024) Artificial Intelligence and Privacy. *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4713111.

Steinert, Steffen et al. (2019). Doing Things with Thoughts: Brain-Computer Interfaces and Disembodied Agency. *Philosophy & Technology*. v. 32, nº 3, p. 457–82. https://link.springer.com/article/10.1007/s13347-018-0308-4

Swan, Melanie (2013). The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data* 1, nº 2, p. 85–99. https://pubmed.ncbi.nlm.nih.gov/27442063/

Taylor, Charles (1985). *Human Agency and Language*. Cambridge: Cambridge University Press. https://philpapers.org/rec/TAYHAA

Taft, William Howard 1(927). Supreme Court Of The United States. U.S. Reports: Olmstead v. United States, 277 U.S. 438. Retrieved from the Library of Congress. https://www.loc.gov/item/usrep277438/.,

Tribe, Laurence H. (1978). *American Constitutional Law*. Brooklin: The Foundation Press, Inc.

UN – UNITED NATIONS (1966). *International Covenant on Civil and Political Rights*. December 16, 1966. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

UNESCO (203). Unveiling the Neurotechnology Landscape - Scientific Advancements Innovations and Major Trends. https://unesdoc.unesco.org/ark:/48223/pf0000386137

Van Cleynenbreugel, Pieter. (2018). Typologies of solidarity in EU law: a non-shifting landscape in the wake of economic crises. BIONDI, Andrea; DAGILYTÈ, Eglè; KÜÇÜK, Esin. *Solidarity in EU Law - Legal Principle in the Making* (Cheltenham, UK: Edward Elgar Publishing Limited. p. 13–37.

Yuste, Rafael. *et al*., (2017). Four ethical priorities for neurotechnologies and AI. *Nature.* v. 551, n. 7679, p. 159–63. https://www.nature.com/articles/551159a

— III —

LEGAL REGULATION OF NEUROTECHNOLOGY

— 7 —

Governing neurotechnology globally: between norms fragmentation and coherence

TSUNG-LING LEE
Taipei Medical University
tl265265@tmu.edu.tw

Introduction

Neurotechnology is attracting an increasing international attention as it has the potential to radically change the way cognitive functioning, wellbeing, and behaviours are understood (Hanslmayr, 2024). The field's unprecedented advancement has sparked considerable enthusiasm and optimism within both scientific and medical communities, particularly regarding its promising clinical applications that hold the promise of restoring vital physical functions in individuals living with paralysis, hearing impairment, or vision loss (UNESCO, 2023). Neurotechnology demonstrates remarkable potential in addressing and treating a broad spectrum of neurological diseases, including but not limited to Parkinson's disease, Alzheimer's disease, and various other forms of dementia—devastating conditions that not only inflict profound suffering on patients and their families but also consistently rank among the leading causes of mortality and disability across the globe (Rao, 2024). The therapeutic reach of neurotechnology extends well beyond traditional neurological disorders, offering promising avenues for treating various mental health conditions, most notably depression—a condition that not only imposes substantial economic burdens on healthcare systems but can also prove severely debilitating for affected individuals, significantly impacting their quality of life and daily functioning (WHI, 2023). For people with neurological disorders, neurotechnology also promises to enhance participation in social, political, and legal processes through the restoration of neurological functioning, thus thereby fostering greater inclusivity across society (The

Neurorights Foundation, 2025). Additionally, in the military sphere, neurotechnology applications have expanded to include sophisticated neural prostheses designed to enhance soldier performance and capabilities, representing a significant advancement in human-machine integration (Ienca; Joterrand, Elge, 2018).

Broadly speaking, neurotechnology functions in three different ways—through "reading," "writing," or both (Kosal; Putney, 2023). Reading technology captures and interprets the brain's electrical activity, while writing technology stimulates electrical activity in neurons. Neurotechnology can be further divided into invasive or non-invasive devices (Kosal; Putney, 2023), where invasive devices require surgical procedures for intracranial implantation. Non-invasive neurotechnology measures brain activities through electroencephalography (EEG) or functional magnetic resonance imaging (fMRI). Currently, only non-invasive neurotechnology is available for commercial use.

While there is no agreed definition on neurotechnology, the OECD defines as "devices and procedures that are used to access, monitor, investigate, assess, manipulate, and emulate the structure and function of neural systems." (OECD, 2025) UNESCO's draft recommendation on the ethics of neurotechnology ethics offers a broader view, describing it as devices, systems, and procedures—both hardware and software—that "directly access, monitor, analyze, predict or modulate the nervous system to understand, influence, restore, or anticipate its structure, activity, function, or intentions (speech, motor)" (UNESCO, 2023). The UNESCO definition also encompasses the convergence of these technologies with neuroscience, engineering, and computing.

One notable application of neurotechnology concerns with brain-machine interference (BMI). Promising therapeutic purposes where individuals affected by Amyotrophic Lateral Sclerosis (ALS), cerebral palsy, stroke or spinal cord can communicate with computers to restore paralysis. Commercial applications of BMI are on the horizon, and some of these applications are marketed to enhance mental focus, or for recreational "neurogaming" or as mind-operated smart phones. However, it should be noted that despite the rapid progress of neurotechnology, it does not provide accurate result in decoding mental through, where mind-reading or steering remain unachievable (UN – Human Rights Council, 2024).

With these promising clinical applications in revolutionize healthcare and tremendous economic potentials, neurotechnology also raises complex ethical, legal, and societal concerns (UN – Human Rights Council, 2024). These include questions about brain data privacy and autonomy, how to regulate direct-to-consumer devices as they merge with artificial intelligence and engineering in an increasingly digitalized era, which are disrupting traditional regulatory boundaries between clinical and non-clinical settings, between medical therapies and consumer markets. As neurotechnology advances, how to protect individuals from cognitive manipulation for commercial or political purposes is also drawing attention (OECD, 2025). Additionally, even after these technologies prove their safety and efficacy through rigorous testing and clinical trials, their substantial development costs and initially limited availability could exacerbate existing social inequalities in healthcare access, risk widening existing social inequalities in access and use to these potentially life-changing innovations. As neurotechnology pose potential threats to privacy, human

autonomy and exacerbate inequalities—all with significant consequences for states and societies, global governance of neurotechnology is attracting a considerable attention at the international level with various international organizations issuing guidelines and recommendations in attempt to guide and steer global policies on neurotechnology.

Three key aspects of neurotechnology help explain why neurotechnology increasingly attracts international governance initiatives.

First, neurotechnology create externalities that are not confined to national borders and whose regulations requires international cooperation. Second, the development of neurotechnology is a cross-border process involving transnational actors—these private actors, multinational technological firms in particular, are shaping and influencing the types of consumer products available on the markets. Third, as neuroscience and engineering converge in the digital age, the boundaries between clinical and non-clinical uses, as well as between medical and consumer applications, become increasingly blurred. Existing domestic and international regulatory frameworks grapple to adequately address these emerging challenges which require a common regulatory to facilitate international commerce of these products.

The global governance of neurotechnology therefore requires concerted efforts at multiple governance levels, sustained cooperation, shared responsibility and accountabilities across multiple sectors and stakeholders. In the past, states have turned to international organizations to solve coordination challenge, promote public goods, or facilitating international commerce. Thus unsurprisingly, the rapid advancement of neurotechnology have prompted several international governance initiatives to manage the externalities of neurotechnology, including the United Nations Human Rights Council, which focuses on human rights implications of neurotechnology (UN – Human Rights Council, 2025), the United Nations Educational, Scientific and Culture Organization (UNESCO) is considering adopting a standard-setting normative instrument on the ethics of neurotechnology (UNESCO, 2024), and the Organization for Economic Co-operation (OCED) became the first international organization that have promulgate nine principles to encourage responsible innovations in neurotechnology (OECD, 2025).

Governance initiatives extend beyond the international level, with significant efforts taking place at both regional and country levels parallel to international initiatives.

At the regional level, the Organization of American States, the Inter-American Declaration of Principles Regarding Neuroscience, Neurotechnologies, and Human Rights (2023) underscores the need to have a clear human rights-based approach. Similarly, the Council of Europe is considering a strategic action plan on technologies in biomedicine (UN – Human Rights Council, 2025, par. 57).

Similarly, at country level, in anticipation of potential human rights implications stemming from neurotechnology, several countries have taken proactive measures. Chile amended its Constitution to introduce rights to mental privacy and integrity (UNESCO Courier, 2025). Similarly, Brazil strengthened its national legislation to enhance the protection of mental data and consent requirements (UN – Human Rights Council, 2025. France has prohibited any activities related to modifying cerebral activities that could pose actual or perceived dangers to human health (UN – Human

Rights Council, 2025). Spain introduced digital rights through a non-binding charter (UN – Human Rights Council, 2025), while China adopted specific guidelines on research ethics for brain-computer interfaces (UN – Human Rights Council, 2025). The United States has focused on safety, security, and consumer rights (UN – Human Rights Council, 2025).

It is important to emphasize that despite rapid technological advances, current neurotechnology capabilities have significant limitations. The state-of-the-art technology cannot accurately capture, interpret, or visualize human thoughts with any meaningful precision (UN – Human Rights Council, 2025). Nevertheless, because neurotechnology represents a potentially transformative development in human-machine interaction and raises novel questions about cognitive privacy and autonomy, it has been deemed as a socially disruptive technology. This classification has catalyzed extensive discussions and debates across multiple disciplines regarding its potential benefits—such as medical applications and assistive technologies—as well as possible risks related to privacy, consent, and human rights (UN – Human Rights Council, 2025). The anticipatory nature of these discussions reflects both the technology's promise and the importance of proactive governance frameworks in today's multipolar world.

My goal in this chapter is descriptive and analytic, to highlight how norms fragmentation occurs as various international organizations with partially overlapping mandates introduce guidance documents and principles driven by varying concerns and interests. In so doing, the paper addresses the current critical gap in the literature. Additionally, I propose solutions for enhancing institutional coherence in neurotechnology governance and highlight the potential norm fragmentation in the global neurotechnology landscape as part of a future research agenda.

The chapter proceeds as follows. Part I begins with an examination of the global landscape of neurotechnology governance, exploring the various regulatory frameworks and policy approaches that have emerged across different jurisdictions. Part II delves into an analysis of the key international organizations that play crucial roles in shaping neurotechnology governance, with particular attention to their overlapping mandates and spheres of influence in this rapidly evolving field. Part III then examines the implications and potential risks of norm fragmentation in the governance landscape, highlighting specific challenges that arise from uncoordinated regulatory approaches. Part IV presents potential solutions and strategies for addressing these governance challenges and identifies a potential international organization to act as the coordinating agency at the international level. Finally, Part V synthesizes the key findings and insights from the previous sections to offer concluding reflections on the future of neurotechnology governance at the global level.

1. Neurotechnology and the Multiplicity of International Organizations Responses

As neurotechnology continues its rapid advancement, the landscape of global governance is evolving into an intricate and multifaceted network of laws, regulations, rules, and guidelines that spans jurisdictions and regulatory domains. This increasing complexity arises from the diverse nature of neurotechnology applications, each

falling under different legal and regulatory frameworks that have been established in time with specific purposes, operational rules, and distinct membership structures. The resulting governance architecture necessarily takes on a decentralized character, functioning across multiple jurisdictional layers with various regulatory frameworks that are tailored to address the particular applications and use cases of different neurotechnology innovations (O'Shaughnessy, 2023). In this complex ecosystem, both state and non-state actors bear significant responsibilities for ensuring compliance with established standards and regulations, with non-state actors—including transnational commercial actors, research institutions, and civil society organizations—emerging as particularly influential forces in developing and implementing global standards and best practices that shape the future of neurotechnology governance.

This decentralized governance structure mirrors the fundamental anarchic nature of international law, characterized by the notable absence of a centralized legal authority with comprehensive enforcement powers. Within the dynamic and rapidly evolving field of neurotechnology, this fragmentation becomes particularly pronounced and consequential—no single international organization possesses the comprehensive oversight authority necessary to govern and monitor the technology's increasingly diverse applications and far-reaching implications. Instead, the complex governance landscape means the authorities to govern is distributed across multiple specialized bodies, each tasked with overseeing distinct yet interconnected aspects of neurotechnology regulation: the UN Human Rights Council addresses critical human rights implications and potential violations, UNESCO actively promotes ethical scientific advancement with neurotechnology, the Organisation for Economic Co-operation and Development (OECD) facilitates free market development and responsible innovation, and the World Health Organization (WHO) is tasked with the harmonization of therapeutic and clinical applications to ensure safety and efficacy.

The growing international attention to neurotechnology's societal implications has also catalyzed a diverse array of voices, each driven by distinct concerns. These include calls for treaties, advocacy for the creation of "neurorights," (The Neurorights Foundation, 2025) the reinterpretation of existing human rights to address emerging concerns (The Neurorights Foundation, 2025, and the establishment of general principles for responsible innovation at the international level (OECD, 2025). The following section will describe the respective roles and mandates of pertinent international organizations involved in neurotechnology.

These governance initiatives and various legislative proposals—while welcoming—inevitably creates a multifaceted regulatory environment where neurotechnology encounters varying standards and guidelines, which differ substantially based on its specific implementation context and intended purpose, while also needing to align with existing domestic regulations and legal frameworks. While this decentralized governance structure offers valuable advantages in terms of specialized expertise and adaptive flexibility to address emerging challenges, it simultaneously presents significant barriers to achieving coherent international standards and implementing well-coordinated approaches across the various domains of neurotechnology development and deployment, potentially leading to regulatory gaps and inconsistencies.

The ensuing part describes the three key institutions in influencing global neurotechnology policies.

1.1. United Nations and the human rights council

At the international level, the United Nations system has emerged from the ashes of World War II as the preeminent international organization for States, notable for its universal state membership and global reach. The UN's fundamental purpose was to facilitate international cooperation through legal commitments—instead of power and politics—and foster peaceful relations among nations (Slaughter, 1994). The UN system operates on several core principles that underpin international relations: sovereign equality, ensuring all member states have equal standing regardless of size or power; sovereign autonomy, protected through self-determination and non-intervention in domestic affairs; and an universal commitments to the international rule of law, where disputes are resolved through agreed rules and processes rather than political resolution of dispute (Slaughter, 1994).

Within the UN system, the UN Human Rights Council is the main intergovernmental body responsible for strengthening and protecting human rights, addressing violations, and making recommendations to safeguard human rights worldwide. Established in 2006 by the General Assembly, the Council consists of 47 Member States and works through international forums with UN officials, mandated experts (special rapporteurs), states, and civil society. The Human Rights Council adopts resolutions and decisions that express the international community's will, serving as a strong political signal for states to remedy human rights violations (UN – HRC, 2025). In October 2022, the Human Rights Council adopted a Resolution on neurotechnology and human rights (UN - Human Rights Council, 2022b). The Resolution notes the potential benefits and threats to human rights, while acknowledging that the impacts, opportunities, and challenges of neurotechnology are not fully understood, and emphasizing the need to analyze the human rights implications in a "coherent, holistic, inclusive and comprehensive manner" (UN - Human Rights Council, 2022b). As a result, the Council requested its Advisory Committee—which serves as the Council's "think tank" and comprises members selected from member states to advise on thematic human rights issues—to prepare a report on neurotechnology's impacts on human rights for presentation at the Council's fifty-seventh session.

The Advisory Committee report notes that neurotechnology's rapid growth is shifting toward the direct-to-consumer market—a largely unregulated sector marked by swift expansion. Some countries lack the necessary technical expertise, capacity, and adequate oversight bodies and regulations. The Advisory Committee cautions that commercial actors may exploit these gaps, prioritizing profits over ethical and human rights considerations (UN - Human Rights Council, 2022a, par. 15). Further concern relates to the risk-benefit assessment for direct consumer products that aims to provide "cognitive enhancement" of healthy person, which differ significantly from assessments for medical use. Accordingly, the Advisory Committee underscores that anticipation is key as part of preparedness to ensure human rights violations do not occur because of actions and misuses. Any framework governing neurotechnology must be flexible enough to provide rapid responses that safeguard human dignity, freedom of thought, privacy, personal integrity, and health (UN - Human Rights Council, 2022a, par. 15). The Advisory Committee therefore underscores the principle of precaution should govern policymaking on neurotechnology (UN - Human Rights Council, 2022a, par. 54).

With various international, regional and countries initiatives emerging, the Advisory Committee therefore cautions that without a minimal level of consistency and coordination amongst the varying approaches, fragmentation of norms is likely. Further, the Advisory Committee is of opinion that "states cannot be left to respond alone to problems as they arise; they may need support as expertise and adequate institutional capacities are lacking" (UN - Human Rights Council, 2022a, par. 60).

The Advisory Committee believes that articulating a human rights framework can help mainstreaming a human rights approach could create cohesion among national responses, guide national and international human rights bodies, provide guidance to public authorities and a tool for the private sector through a soft-law instrument where the Human Rights Council should take the lead in cooperation with the Office of the United Nations High Commissioner for Human Rights. The Advisory Committee is of view that a set of guiding principles on the application of the human rights framework to neurotechnology should be developed (UN - Human Rights Council, 2022a, par. 60).

In sum, the UN Human Rights Council's approach towards neurotechnology can be characterized as a precautionary and anticipatory one, driven by concerns that its development will outpace regulators' and human rights bodies' ability to understand and address its impacts in a timely manner.

1.2. The United Nations Educational, Scientific and Cultural Organization

Established in 1945, UNESCO has the primary constitutional task of promoting collaboration among nations in education, science, and culture. The Organization, whose key responsibility is to contribute to peace and security through global collaboration in the areas within its competence, has 194 member states and 12 Associate Members (UNESCO, 2025). It is one of the largest specialized agencies in the United Nations (UN) and the sole UN agency mandated for human and social science (Langlois, 2017). UNESCO's Constitution provides the Organization with a broad mandate that across three fields—education, science, and culture—this means that the remit of the Organization can overlap with other UN agencies as no other UN specialized agency enjoys such a broad mandate (Romano; Boggio, 2014). Established in the aftermath of World War II, founders of the Organization held a firm belief in education, science, and culture as a foundational driver for peace. This belief is reflected in the purpose of the Organization. Article I, paragraph 1 provides that UNESCO is "to contribute to peace and security by promoting collaboration among the nations through education, science, and culture in order to further universal respect for human rights and fundamental freedoms" (UNESCO, 1945). The Organization has an active record of promoting science and human rights within UNESCO's competence, running operational activities since 1948.

To fulfill its institutional mission, UNESCO's founders granted the Organization legal authority to adopt conventions and recommendations under Article IV, Paragraph 4 of its Constitution. With universal membership, UNESCO plays an influential and active role in bioethics. The Organization has established various bioethical standards, including the Universal Declaration on the Human Genome and Human Rights, the International Declaration on Human Genetic Data, the Universal

Declaration on Bioethics and Human Rights, and guidelines on AI. A new recommendation on neurotechnology is set for potential adoption later this year (UNESCO, 2023). With its universal membership, the recommendation on the ethics of neurotechnology could potentially become influential and steer policy directions on neurotechnology.

1.3. Organization for Economic Cooperation and Development

Established in 1961, the Organization for Economic Cooperation and Development (OECD) consists of 38 member countries where the organization is focused on promoting economic growth, prosperity, and sustainable development. It succeeded the Organization for European Economic Cooperation, which was created to implement the Marshall Plan after World War II. The OECD operates through diplomatic means where nearly 169 of its recommendations are adopted by consensus, function as normative framework for a larger array of policies (Winickoff, 2021). OECD serves as a hub for data and analysis, facilitates evidence-based policymaking, and sets international standards. Despite its relatively small membership, these member countries account for three-fifths of the world's total GDP while representing only 18% of the global population (Winickoff, 2021). The OECD wields substantial economic influence through its member countries' voluntary adoption of its recommendations.

OCED defines its recommendations as "not legally binding but practice accords them great moral force as representing the political will of Adherents. There is an expectation that Adherents will do their utmost to fully implement a Recommendation" (OECD, 2025). Soft law carries significant weight within the OECD and represents a strong political commitment by member countries—so much so that some of the organization's most influential standards are not legally binding. For example, the Polluter Pays Principle first appeared internationally in an OECD Recommendation on Guiding Principles concerning International Economic Aspects of Environmental Policies, twenty years before its inclusion in the Rio Declaration on Environment and Development (Bonucci, 2011). Similarly, the OECD's development on the recommendation of the Council of Artificial Intelligence has shaped the G7 Agenda on AI and helped create an AI Policy Observatory (Winickoff, 2021).

For the governance of neurotechnology, the OCED takes a "responsible innovations" approach (Kosal; Putney, 2023) which seeks to address the "Collingridge Dilemma"—regulating to early could stifle innovation while delaying regulation until the technology is widely adopted could allow problematic practices to become deeply embedded in society. This approach emphasizes proactive governance through three key strategies: anticipating potential challenges and risks before they materialize, actively steering development toward optimal societal outcomes, and ensuring meaningful inclusion of diverse stakeholders throughout the innovation process. Unlike the Advisory Committee stance on commercial actors, the OCED responsible innovations approach advocates for co-creation of norms between industry and individuals. While the Advisory Committee underscores the principle of precaution, the OCED recommendation encourages regulatory experimentation in the form of testbeds, sand boxes, new technology assessment methods and foresight strategies (Kosal; Putney, 2023).

2. Fragmentation

Neurotechnology encompasses a diverse range of applications that intersect with numerous international regulatory frameworks, including but not limited to trade regulations, military protocols, human rights laws, and research ethics guidelines. The landscape of stakeholders engaged in its global governance is remarkably extensive, encompassing public and private research institutions, military organizations, insurance providers, clinical regulatory authorities, and consumer protection bodies, each bringing distinct perspectives and requirements to the governance dialogue. Given the intricate nature of these overlapping jurisdictions and interests, it becomes evident that the global governance of neurotechnology cannot be effectively contained or managed within the confines of any single institutional framework or regulatory regime. Rather, it necessitates coordination among multiple actors, regulatory processes, and institutions, each operating with its own distinct set of interests, strategic objectives, and foundational values.

This multiplicity of governance arrangements has given rise to an increasingly sophisticated global ecosystem on neurotechnology, creating potential fragmentation of norms. Fragmentation of international law refers to both the process and outcome of multiple institutions with overlapping mandates operating in a non-hierarchical legal order (Kreuder-Sonnen; Zürn, 2020). This fragmentation is beginning to manifest in the field of neurotechnology, where various applications fall under different international regimes with interrelated jurisdictions. When international organizations have overlapping mandates, it can create normative and institutional conflicts—a phenomenon first analyzed in Professor Koskenniemi's influential work on institutional politics at the international level (Koskenniemi, 2006).

As noted above, three key institutions have begun addressing emerging neurotechnology challenges, each with distinct objectives: the OECD promotes market efficiency and responsible innovations, UNESCO facilitates ethical scientific development, and UN human rights bodies safeguard fundamental rights. While the mandates of these organizations differ considerably, they face two major common governance challenges: their recommendations lack enforcement mechanisms and serve only as guidance, and early indicators suggest fragmented international rules may emerge. Second, the OECD, UNESCO, and UN human rights bodies have historically operated in isolation, making it likely that neurotechnology governance will develop without coordinated policy responses or a coherent approach. This siloed approach is at risk of creating disconnected responses—and when multiple conflicting rules exist, it effectively means no rules exist at all. This piecemeal pattern extends to regional and national levels, where the Human Rights Council Advisory Committee on neurotechnology has noted that uneven responses could create inconsistent human rights protections (UN – Human Rights Council, 2024). The Committee warns that without proper consistency and coordination amongst governance efforts, these isolated responses will be inadequate to tackle the social, ethical and human rights concerns arising from neurotechnology.

Fragmentation occurs along two dimensions: substantive norms and international authority to govern. Substantive norms, which include sources of international law, encompass both hard and soft law. While no binding treaty currently governs neurotechnology specifically, various soft law instruments—guidelines, recommen-

dations, and opinions—are emerging and shaping the governance landscape. Multiple international organizations with overlapping mandates are involved in governing neurotechnology, varying by the specific application and deployment of the technology. Since there is no centralized authority in the international system, the authority to govern—or the distribution of power—is shared (often unequally) among numerous international institutions that produce, interpret, and apply these norms according to their established mandates.

This multiplicity of governance arrangements has given rise to an increasingly sophisticated and interconnected global ecosystem that continues to evolve as neurotechnology advances. However, this complexity can extend far beyond the mere potential for normative conflicts and institutional competition between regulatory bodies. The partial overlap in mandates among these various organizations involved in neurotechnology governance could result in substantial practical challenges that affect the field's development and oversight.

These challenges can manifest in two ways: first, when multiple organizations independently pursue similar initiatives, it might result in duplication of resources; second, through the creation of redundant and sometimes conflicting regulatory frameworks that create confusion for stakeholders. These overlapping jurisdictions and competing priorities can significantly impede the development of coherent global oversight mechanisms, potentially creating governance blind spots. The resulting fragmentation not only complicates compliance for organizations operating across multiple jurisdictions but also poses risks to the consistent application of ethical standards and safety protocols in various neurotechnology development and deployment.

Additionally, soft law instruments that lack robust enforcement mechanisms face significant challenges in achieving their intended impact. Commercial entities, driven by market pressures and profit motives, frequently default to implementing only the minimum required standards, rather than striving to meet higher, yet reasonably achievable benchmarks that would better serve the public interest. This tendency toward minimal compliance undermines the effectiveness of voluntary guidelines and recommendations. Furthermore, the proliferation of overlapping and sometimes contradictory regulatory frameworks creates a particularly problematic scenario - when multiple conflicting rules exist simultaneously, the practical result is often paralysis, where the abundance of competing standards effectively nullifies their collective authority and renders meaningful compliance virtually impossible.

Moreover, the diversity and complexity of regulatory frameworks applicable to neurotechnology could create numerous opportunities for strategic "forum shopping," where both governmental bodies and commercial entities may selectively choose policies, standards, and jurisdictions that align with their interests and objectives, rather than prioritizing broader public welfare and societal benefits. This practice becomes especially concerning when multiple institutions possess overlapping jurisdictional mandates and maintain competing or inconsistent standards. Such regulatory fragmentation can have far-reaching consequences, potentially undermining the credibility, effectiveness, and overall authority of all institutions involved in the governance framework. The resulting regulatory uncertainty may not only com-

plicate compliance efforts but also create gaps in oversight that could compromise ethical, human rights and regulatory considerations.

2.1. Fragmentation as part of international legal system

On the other hand, the fragmented nature of global neurotechnology governance appears inevitable, given both the technology's inherent complexity and the decentralized nature of the international legal system. This fragmentation largely stems from neurotechnology's applications crossing multiple regulatory domains—trade, military, human rights, and health. The field's diverse stakeholders reflect this plurality, ranging from research organizations and military forces to insurance companies and regulatory bodies that oversee both clinical and consumer applications. Given this complexity, neurotechnology's global governance cannot be effectively managed through a single institution or regulatory regime.

Thus, fragmentation of norms in the global governance of neurotechnology is arguably inevitable since no single centralized authority can coordinate and enforce international norms uniformly across different domains and jurisdictions. As the world has grown in complexity and evolved from monopolar into a multipolar system, the traditional global legal order has evolved from a hegemonic structure to a more diverse pluralistic framework. This transformation has created a multiplicity of legal frameworks, regulatory bodies, and decision-making centres, resulting in a more nuanced but challenging environment for international governance and cooperation.

While fragmentation might be an inevitable consequence of different and diverse neurotechnology applications, coherence of norms amongst the relevant regulatory regime might remain desirable. Coherence creates predictability and brings equality to international legal system, as it demands that like cases to be treated alike, as argued by theorist Niklas Luhmann (1997). While fragmentation is inevitable, global governance of neurotechnology demands coordination among multiple actors, processes, and institutions—each with distinct interests, agendas, and values.

Additionally, as neurotechnology continues to advance at a rapid pace, the regulatory capability to thoroughly evaluate and assess clinical applications has become not only essential for ensuring patient safety and therapeutic efficacy, but also strategically advantageous for securing market access and commercial viability. These regulatory capabilities and institutional frameworks exhibit considerable variation across different regions and jurisdictions worldwide—while some states have developed robust infrastructure, accumulated substantial resources, and cultivated the necessary technical expertise to effectively address emerging regulatory challenges, others face significant constraints in terms of institutional capacity, technical knowledge, and regulatory experience, necessitating additional support, targeted capacity building initiatives, and international cooperation to develop their regulatory frameworks.

While global concerted efforts toward neurotechnology governance would be ideal for market efficiency and unified human rights protection, two fundamental factors suggest that pluralistic approaches at the national level are more likely. First, sovereignty—the locus of decision-making about neurotechnology remains with individual states. Second, states vary greatly in their capabilities—while some possess

the resources and technical expertise to address regulatory challenges, others require additional institutional support and capacity building (UN – Human Rights Council, 2024, par. 60).[1] These disparities in regulatory capability and resources indicate that imposing unified standards could for create disproportionate burdens for countries that currently lack the necessary technical resources, regulatory frameworks, and specialized expertise required to effectively govern neurotechnology development and implementation.

Accordingly, this fragmentation might not necessarily be a bad thing, instead, it can serve as a catalyst for creating a more dynamic, responsible and inclusive pluralistic global legal structure. A pluralistic legal order might be desirable as it might better accommodate diversity of views and values. Additionally, institutional competitions might spur international organizations to be more agile and anticipatory of negative externalities that stem from neuroethologies. While the guidelines, reports and the potential UNESCO recommendations are all soft law in nature—which mean they are not legally binding—this flexibility enable countries to tailor these recommendations from international organizations according to their specific country needs and contexts.

Moreover, fragmentation of norms is not isolated to neurotechnology. In fact, fragmentation is commonly observed in the international legal system, where the phenomenon has become the norms, rather than the exception. One prime example is the global governance of artificial intelligence, where multiple rules co-exist with various of soft law from constellation of international organizations (Tallberg; Erman; Furendal; Geith; Klamberg; Lundgren, 2023). These soft law instruments also co-exist with regional governance efforts such as the EU AI Act.

3. Managing fragmentation – towards coherence of norms

As neurotechnology blurs traditional regulatory boundaries, governance authority has become increasingly fragmented among international organizations with overlapping mandates and interrelated interests. While such fragmentation is common across sectors, early indicators suggest that neurotechnology's global governance may mirror AI governance's decentralized nature. However, unlike AI, where norms have evolved alongside the technology itself, neurotechnology as a field remains at an early stage, making it difficult to assess how emerging international norms will shape the development of neurotechnology. However, since fragmentation of international law is the norm rather than the exception, the question is how to manage fragmentation amongst the multiplicity of institutions and mitigate the potential competing or conflicting of norms thus become the central question.

The coordination of norms among international institutions in the neurotechnology space can be achieved through two primary approaches: either through the establishment of a centralized regulatory agency or through enhanced cooperation and coordination between existing relevant agencies. While the current state of inter-

[1] "60. Without a minimum degree of consistency and coordination among such approaches, isolated and fragmented responses are insufficient. Given the particularities of the neurotechnologies industry, ensuring similar levels of protection and coherence in implementation is key. States cannot be left to respond alone to problems as they arise; they may need support as expertise and adequate institutional capacities are lacking".

national relations and the complex nature of global governance make the creation of a centralized agency for neurotechnology particularly challenging, the strengthening of cooperation among key institutional agencies emerges as a more pragmatic and achievable solution. The climate governance regime provides an instructive example of successful multi-institutional coordination, where governance responsibilities are effectively distributed among various UN agencies and actions are implemented through a polycentric order utilizing multilevel approaches that engage stakeholders at global, regional, and national levels.

In the current landscape, the UN Human Rights Council, UNESCO, and the OECD serve as primary norm-generating sites for neurotechnology governance, functioning as constituent elements within what can be characterized as a "weak regime complex" (Roberts; Hine; Taddeo; Floridi, 2024) - a network of loosely connected institutions operating within a pluralistic governance structure. The transition from this weak regime complex to a more robust and effective governance framework would necessitate several key developments: significantly enhanced institutional cooperation, the establishment of formal partnerships, and the implementation of clear conflict resolution mechanisms to address potential jurisdictional overlaps and policy conflicts.

Moreover, achieving improved policy coordination requires a substantial enhancement in democratic legitimacy across these three institutions, particularly in their approach to neurotechnology governance. This necessitates creating meaningful opportunities for diverse stakeholder input on critical neurotechnology issues and policy decisions. Such democratization would involve systematically broadening participation to include representatives from patient advocacy groups, industry stakeholders, research institutions, and civil society organizations, while simultaneously working to identify and align common institutional objectives and governance priorities. Given its substantial economic influence and established track record in facilitating international cooperation on complex technological issues, the OECD appears particularly well-positioned to take a leadership role in coordinating these collaborative efforts and fostering more cohesive global governance of neurotechnology.

However, the OECD's limited membership base presents a significant challenge that could potentially undermine its political legitimacy as a coordinating agency in the global context. Nonetheless, as shown in the climate regime, changes and cementing institutional partnerships are incremental processes. The OECD as an expert body could leverage its technical role in support of other international institutions and inform global neurotechnology policies by disseminating research and policy ideas, developing s indicators to measure country-level preparedness for addressing neurotechnology's ethical, social and regulatory impacts, and the establishing a dedicated neurotechnology policy observatory to monitor global developments in neurotechnology and facilitate knowledge sharing across jurisdictions.

Conclusion

Neurotechnology demonstrates remarkable potential for advancing human capabilities and medical treatments, but its development and implementation necessitate thorough consideration of human rights implications and ethical boundaries. This

technological frontier requires ongoing, in-depth dialogue among diverse stakeholders, including researchers, funders, policymakers, transnational companies, lawyers, legal scholars, clinicians, patient groups, ethicists, and affected communities. To effectively address evidence-based concerns and potential risks, meaningful participation from all relevant parties must be sustained throughout the development process, ensuring that both scientific progress and human rights protections advance in tandem. International organizations play crucial roles in developing norms that govern neurotechnology. However, with multiple organizations driven by diverse mandates and values, fragmentation of norms is likely. This fragmentation could produce normative conflicts, create regulatory confusion, as well as resource duplication. Yet, as demonstrated in this chapter, coherence can be achieved through coordinated international efforts, with the OECD well-positioned to take on this coordination role.

References

Bonucci, Nicola. (2011) The Oecd at Fifty: Some Observations on the Evolving Nature of an International Organization, The Georg Washington. International Law Review, v. 43, n. 2, p. 239-247, 2011. https://heinonline.org/HOL/LandingPage?handle=hein.journals/gwilr43&div=12&id=&page=

Hanslmayr, Simon. (2024). The Promises and Challenges of Neurotechnology to Improve Human Health and Cognition. *PLOS Biology*, October 30. 2024. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002903

Ienca, Marcello; Jotterand, Fabrice; Elge, Bernice. (2018). From Healthcare to Warfare and Reverse: How Should We Regulate Dual-Use Neurotechnology? *Neuron*, v. 97, n. 2, p. 269-274, Jan 17, 2018. https://pubmed.ncbi.nlm.nih.gov/29346750/

Kosal, Margaret; Putney, Joy. (2023). Neurotechnology and International Security: Predicting Commercial and Military Adoption of Brain-Computer Interfaces (BCIs) in the United States and China. *Politics and the Life Sciences*, v. 42, n. 1, p. 81-103, April, 2023. https://pubmed.ncbi.nlm.nih.gov/37140225/

Koskenniemi, Martti (2006). Report of the Study Group of the International Law Commission. Fragmentation of International Law: difficulties arising from the diversification and expansion oflinternational Law, document A/CN.4/L.682 and Add.1. April, 13, 2006. https://legal.un.org/ilc/documentation/english/a_cn4_l682.pdf

Kreuder-Sonnen, christian; Zürn, Michael (2020). After Fragmentation: Norm Collisions, Interface Conflicts, and Conflict Management. Global Constitutionalism, v. 9, n. 2, p. 241, 241-267. https://www.researchgate.net/publication/339461448_After_Fragmentation_Norm_Collisions_Interface_Conflicts_and_Conflict_Management

Langlois, Adèle. (2017). The Global Governance of Human Cloning: The Case of UNESCO. *Palgrave Communications*, March 21, 2017. https://www.semanticscholar.org/paper/The-global-governance-of-human-cloning%3A-the-case-of-Langlois/dcca7b2a43043f9e0d0787b98d0f02e0d72af089

Luhmann, Niklas (1997). *Die Gesellschaft der Gesellschaft*, 2 vols. Frankfurt am Main: Suhrkamp.

O'Shaughnessy, Matthew R. et al. (2023) *Neuroethics Guid*ance Documents: Principles, Analysis, and Implementation Strategies" 10 Journal of Law and the Biosciences lsad025.

OECD - The Organisation for Economic Co-operation and Development. *Legal Instruments*. (2025) https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Rao, Naveen (2024). No Moonshots with Paper Planes: Alzheimer's Trials Need Neurotech. Forbes, January 30, 2024. https://www.forbes.com/sites/naveenrao/2024/01/30/no-moonshots-with-paper-planes-alzheimers-trials-need-neurotech/

Roberts, Huw; Hine, Emmie; Taddeo, Mariarosaria; Floridi, Luciano. (2024). Global AI governance: barriers and pathways forward. *International Affairs*, v. 100, n. 3, p. 1275–1286, May 2024. https://academic.oup.com/ia/article/100/3/1275/7641064

Romano, Cesare PR; Boggio, Andrea. (2024). *UNESCO and the Right to Science*. Oxford Academic. https://academic.oup.com/book/58126/chapter-abstract/479822273?redirectedFrom=fulltext

Slaughter, Anne-Marie. (1994). The Liberal Agenda for Peace: International Relations Theory and the Future of the United Nations, *Transnational Law and Contemporary Problems*, v. 4, p. 377- *386,*1994.

Tallberg, jonas; Erman, eva; Furendal, markus; Geith, Johannes; Klamberg, Mark; LUNDGREN, Magnus. (2023). The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research. *International Studies Review*, v. 25, n. 3, September 2023. https://academic.oup.com/isr/article/25/3/viad040/7259354

THE NEURORIGHTS FOUNDATION (2025). https://neurorightsfoundation.org

UN – HUMAN RIGHTS COUNCIL (2025). Home. https://www.ohchr.org/en/hrbodies/hrc/home

UN – HUMAN RIGHTS COUNCIL (2024). Report of the Human Rights Council Advisory Committee, Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights, Fifty-seventh session 9, September9 October 2024 https://docs.un.org/en/A/HRC/57/61

UN – HUMAN RIGHTS COUNCIL. (2022). *New and Emerging Technologies and Human Rights at the Heart of New Resolutions Adopted.* October 13, 2022 https://ecnl.org/news/un-hrc-51-new-and-emerging-technologies-and-human-rights-heart-new-resolutions-adopted

UN – HUMAN RIGHTS COUNCIL. (2022a). *Resolution* A/HRC/57/61 Fifty-seventh session 9 September–9 October 2024 Agenda items 3 and 5 https://ecnl.org/sites/default/files/2022-10/Res2.pdf

UN – HUMAN RIGHTS COUNCIL. (2022b*). Resolution A/HRC/51/L.3* Fifty-first session 12 September–7 October 2022. https://ecnl.org/sites/default/files/2022-10/Res2.pdf

UNESCO - The United Nations Educational, Scientific and Cultural Organization. (2023). *The Risks and Challenges of Neurotechnologies for Human Rights* https://unesdoc.unesco.org/ark:/48223/pf0000384185

UNESCO - The United Nations Educational, Scientific and Cultural Organization. (2024). First Draft of a Recommendation on the Ethics of Neurotechnology (Revised Version): Working Document as of 27 August 2024. https://unesdoc.unesco.org/ark:/48223/pf0000391074

UNESCO - The United Nations Educational, Scientific and Cultural Organization. (2023). Toward an international document, November 2023. https://www.unesco.org/en/ethics-neurotech/recommendation

UNESCO - The United Nations Educational, Scientific and Cultural Organization. (1954). *Constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO).* Adopted in London on 16 November 1945, https://unesdoc.unesco.org/ark:/48223/pf0000382500

UNESCO Courier (2022). Chile: Pioneering the Protection of Neurorights. March 21, 2022. https://courier.unesco.org/en/articles/chile-pioneering-protection-neurorights>

WHO – WORLD HEALTH ORGANIZATION (2023). *Depressive Disorder (Depression),* March 31, 2023. https://www.who.int/news-room/fact-sheets/detail/depression

Winickoff, David. (2021)). Neurotechnology at the OECD: The Role of the Private Sector in Governance. Annales des Mines - Réalités industrielles. v. 3, p. 7-11, Août, 2021. https://www.researchgate.net/publication/363354139_Neurotechnology_at_the_OECD_The_role_of_the_private_sector_in_governance

— 8 —

Emerging technologies and protection of human rights: applying the UNGPs in the neurotechnology field

VIVIANE CEOLIN DALLASTA DEL GROSSI
Federal Public Defender
viviane.dallasta@dpu.def.br.

Contents: Introduction; 1. B-Tech Project of UN; 2. Mechanisms to identify and address technosocial impacts and protect Human Rights; 3. State of the art of potential safeguards with immediate practical application; Conclusions; References.

> "Neurotechnology collects intimate personal information that people are often not aware of, including emotions and complex behaviour. The consequences could be dire if these technologies are developed or deployed inappropriately."
> Stephen Almond (2023) *Executive Director of Regulatory Risk* - ICO/UK.

Introduction

This research aims to verify how, in practice, the application of the United Nations Guiding Principles on Business and Human Rights (UNGPs) by technology companies is occurring, particularly by investigating the Human Rights Due Diligence (HRDD) mechanisms used to ensure respect for human rights in the employment of emerging technologies such as neurotechnologies and AI.

In this regard, it aims to identify due diligence mechanisms of potential use by technology companies to mitigate risks and prioritize human rights in practice, as defined in national and international human rights protection documents, while noting the still prevailing regulatory legislative vacuum on the topic.

The work takes into account that AI systems intersecting with neurotechnologies should be seen as "sociotechnical systems" or "technosocial systems," as preferred in the study. This is because the impact of an AI system—regardless of its underlying technology—depends not only on the design of the system but also on how the system is developed and used within a broader environment. This environment includes its data, intended objectives, functionalities and accuracy, the scale of deployment, and the wider organizational, social, and legal context in which it is accessed.

The positive or negative consequences of such systems also depend on the values and behavior of the humans who develop, deploy, and use them, highlighting the importance of ensuring corporate responsibility.

For a better understanding of such impacts, Professor Martyna Kalvaitytė's research (Kalvaitytė, 2022) elucidates the characteristics of the impacts that this work refers to as "technosocial" in the following way:

> The revolutionary nature of changes brought by AI leads to the emergence of new types of impacts, some of which are not covered by existing protections. The impacts that go beyond the individual level include communal and societal levels of impact. These impacts warrant special attention since the approaches to protections against individual negative impacts take precedence, while social impacts tend to be overlooked (Campolo et al. 2017; Griffin 2022; Smuha 2021a). The overlap between individual and societal can take place; however, societal harms have a different scope of impact by affecting societal interests (Smuha 2021a:5). Societal harms warrant attention because they are more difficult to establish since they tend to be caused by repeated interactions that lead to incremental changes, which result in noticeable impacts after longer periods of time than individual harms (Smuha 2021a). The challenges of addressing societal impacts can also be interpreted through the direction of present AI governance approaches as problem-solving instead of problems-finding (Liu and Maas 2021). In the problem-solving governance mode, issues that have already been identified can receive more attention than problems that are new, such as societal AI impacts that challenge boundaries set by existing legal systems (Liu and Maas 2021). As a result, the thesis places special attention on social impacts, given the high stakes of negative impacts against which protections do not exist. (Kalvaitytė, 2022).

Thus, this work uses the term "technosocial impact" to refer to collectively generated impacts with the potential to restructure human coexistence and life in the long term, in ways that are not democratically deliberated. The investigation, prevention, and eventual remediation of these impacts should be established by public policies and legally addressed as protection of diffuse digital rights and interests. They deserve a precautionary approach to potential negative damages, which can generate irreversible consequences, especially when they affect children and adolescents due to their particular stage of human development.

There are still many sectors, including neurotechnologies, and contexts that lack regulation for the preservation of human dignity in interactions within this new scenario. It is evident that this is a crucial moment for imposing safeguards for present and future generations, and for understanding what kind of society we want.

To this end, national and international documents and statements will be presented that highlight the need for debate around digital interactions, including in immersive environments,[1] and that call upon companies, academia, the scientific community, and governments for the importance of regulation. These regulations are seen as extensions of the fundamental rights to privacy, data protection, freedom of thought, and the free development of personality.

Therefore, it is essential to question how to identify and address such impacts. The practical application of the United Nations Guiding Principles on Business and Human Rights (UNGPs) in the technological space plays a crucial role in this field.

The B-Tech Project of the United Nations (UN) follows this approach.

[1] Considering that a significant potential for the application and dissemination of neurotechnologies is precisely in this field of Web 4.0 (UE, 2023). *EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition*. Communication from the Commission to the European Parliament, (2033) the Council, the European economic and social committee and the committee of the regions.

1. B-Tech Project of UN

The United Nations High Commissioner for Human Rights (OHCHR) is developing the B-Tech Project (B-Tech, 2019), which provides reliable guidance and resources for implementing the United Nations Guiding Principles on Business and Human Rights (UNGPs) in the technological space. In 2019, the OHCHR launched the project after consultations with civil society, businesses, states, and other experts regarding the scope of the B-Tech Project.

The B-Tech Project offers guidance and resources to implement the United Nations Guiding Principles on Business and Human Rights (UNGPs) in the technology sector, focusing on four key areas:

Focus 1: Addressing Human Rights Risks in Business Models;

Focus 2: Human Rights Due Diligence and End Use;

Focus 3: Accountability and Remedy;

Focus 4: "A smart mix of measures" - regulatory and policy responses to human rights challenges linked to digital Technologies (B-Tech, 2019).

The B-Tech Project is an innovative initiative by the United Nations to address the urgent need for pragmatic and principled ways to prevent and address human rights harms related to the development of digital technologies and their use by corporate, governmental, and non-governmental actors, including individual users.

In this context, considering Focus 1 (Addressing Human Rights Risks in Business Models) presented above, there is a need to understand the human rights impacts resulting from the business models of companies.

Additionally, there is a need to understand the technosocial impacts of the products and services developed and marketed by technology companies (Big Techs).

2. Mechanisms to identify and address technosocial impacts and protect Human Rights

In 2023, UNESCO published a compilation with viewpoints from various experts to explore the risks of brain technologies, with the aim of investigating whether the existing international legal frameworks are sufficient to protect human rights (UNESCO, 2022).

In the same vein, in June 2023, the UK's data protection authority (Information Commissioner's Office – ICO) published a report summarizing the state of the art of ethical and legal concerns. The report warns that "emerging neurotechnologies risk discriminating against people if these groups are not placed at the center of their development," and predicts that "the use of these technologies will become widespread in the next decade" (ICO, 2023)

Implementing Human Rights Due Diligence (HRDD) mechanisms by companies involves conducting Human Rights Impact Assessments (HRIAs) in technology companies as an embodiment of transparency and practical Human Rights Due Diligence (HRDD) application. This process includes regular publications, periodic reassessments of these mechanisms, and the engagement of independent consultancies as part of their comprehensive "Sustainability" disclosures, encompassing both social and technosocial responsibility. Additionally, there is a need for investors and ven-

ture capital firms to take responsibility for these considerations in their investments, supported by state regulation.

Thus, we can visualize HRDD in the first draft of UNESCO's Recommendation on the Ethics of Neurotechnology on the following topics:

> 124. Member States should conduct impact assessments for all new neurotechnology, including an evaluation of their impact on existing mental health disparities and their consequences. Mental health advocacy groups and population-specific stakeholders should be included in the design of these impact assessments to understand the effects of neurotechnology on various mental health conditions.
>
> [...]
>
> 140. Member States should ensure rigorous benefit-risk assessments in all neurotechnology research involving patients, as standard practice. These assessments must be thoroughly documented, justifying the rationale for the research based on the anticipated benefits outweighing the potential risks. Ethical oversight bodies should review these assessments to maintain high standards of patient safety and well-being. Moreover, ongoing monitoring should be employed to adjust protocols in response to emerging data throughout the research process.
>
> [...]
>
> 141. Member States should ensure research involves strict oversight and close follow-up of all neurotechnology research involving children and adolescents. This oversight is crucial during the developmental phases of childhood to address and mitigate any unforeseen long-term effects. Such research must include comprehensive monitoring protocols and periodic evaluations to ensure the ongoing safety and well-being of young participants, taking into account their unique developmental needs and vulnerabilities.
>
> [...]
>
> 152. Clinical trials should be understood as properly designed interventions with adequately selected human groups in a comparative fashion with other modes of dealing with the specific problem under study. Ethical impact assessments should be transparent and open to the public when appropriate. Assessments should be multidisciplinary, multi-stakeholder, multicultural, pluralistic and inclusive. Assessment and intervention should identify impacts on human rights.
>
> [...]
>
> 208. Member States should strengthen comprehensive consumer protection laws to include clear labelling on commercial neurotechnology products, detailing their effects, limitations, and risks to prevent misleading claims and ensure transparency. This also includes prohibiting practices of "tying" or requiring the disclosure of neural and cognitive biometric data as a condition to access goods or services.
>
> [...]
>
> 218. Member States should develop oversight mechanisms to prevent the use of neurotechnology to drive compulsive or addictive use, particularly if they target neural reward circuits to alter consumption behaviours. Such regulations should mandate clear labeling of risks, enforce game design standards that discourage compulsive use of gaming or digital recreational platforms combined with neurotechnology, to promote healthy, balanced use, especially among children and vulnerable populations.
>
> [...]
>
> 222. Member States should require companies to disclose when AI algorithms, based on neural and cognitive biometric data, are used in making decisions that affect consumers, such as personalized advertising and content recommendations. Transparency should also extend to the algorithms themselves, potentially through mechanisms like algorithmic audits (UNESCO, 2024).

Potential issues that could be applicable in impact assessments on a personal/individual level (ISC, 2023):

> a) Potential effects on people's well-being: Could the system impact (and in what direction) the well-being of individual users (e.g., quality of work, education, social interactions, mental health, identity, environment)?
>
> b) Autonomy: Could the system affect human autonomy by causing excessive dependence on technology by end users?

c) Human Development: Is there an impact on the acquisition of essential skills for human development, such as executive functions, interpersonal skills, changes in attention span affecting learning, personality development, mental health concerns, etc.?

d) Mental Health: Is there a risk of increased anxiety, loneliness, or other mental health issues, or can the technology mitigate these impacts?

e) Human Evolution: Could the technology lead to changes in human evolution?

Additionally, considering the terms set out in the OECD Recommendation on Responsible Innovation in Neurotechnology:

6. Enable the capacity of oversight and advisory bodies to address novel issues in neurotechnology. To this end, relevant actors should:

a) Encourage regulatory agencies, funding bodies, research institutions and/or private actors to respond to opportunities and ethical, legal and social issues raised by advances in brain research and neurotechnology.

b) Encourage research into the ethical, legal and social dimensions of neurotechnology.

c) Promote the further development of ethical guidance and best practices including rigor and reproducibility.

d) Ensure that oversight and advisory bodies possess appropriate multi-disciplinary expertise for constructive technology assessment, horizon scanning, scenario planning, and review of research.

e) Develop institutional capacity and mechanisms of technology appraisal and/or foresight to anticipate and evaluate potential neurotechnology outcomes and pathways (OECD, 2019).

Framework for Anticipatory Governance of Emerging Technologies (OECD, 2024), with the follow Key Concepts:

Emerging Technologies: Technologies characterized by rapid development, evolution, novelty and uncertainty in trajectory and impact. Key examples include new gene editing tools and synthetic biology, neurotechnology, the latest advanced AI technologies, immersive technologies and quantum technologies.

Responsible innovation: Trustworthy technology development (from agenda-setting, to applied research and development and commercialization phases, See Figure 2.2) that is guided by democratic values, responsive to social needs and accountable to society. Responsible innovation depends on effective anticipatory governance (see, e.g. OECD Recommendation on Responsible Innovation in Neurotechnology - OECD, 2019).

The document employs a method of Horizon Scanning as part of anticipatory governance, which requires policymakers to monitor newly emerging technologies that could either pose specific threats or offer socio-economic and sustainability benefits. Horizon scanning involves the regular exploration of new technology developments and signals, identifying which technologies (e.g., neurotechnology) or technosocial issues (such as the digital or quantum divide) are becoming significant areas of interest, viewed either as opportunities or potential concerns. Once an area of interest has been identified through horizon scanning, the next step is to determine whether deeper analysis is necessary to take policy action and what forms of strategic intelligence are required (OECD, 2024).

Additionally, considering the terms set out in the IEEE Neuroethics Framework (IEEE, 2023) that address the ethical, legal, social and cultural implications of neurotechnology, it can be stated that there are already potential human rights due diligence (HRDD) mechanisms applicable to investigate and address the technosocial impacts arising from neurotechnologies.

From this perspective, it is important to note that self-regulation, through the application of the UNGPs, is one of the main tools with real impact available for

policy-making. By itself, it cannot replace the law, but if properly implemented, it is a crucial tool, especially in the following situations (Floridi, 2018):

a) When legislation is not yet in force (for example, in experiments with augmented reality products or neurotechnologies), or when there is no relevant national legislation;

b) When legislation exists but also requires ethical interpretation (for example, in terms of understanding a right to explanation provided by law);

c) Legislation is already in force but also needs some ethical counterbalance: 1) if it is better not to do something, even if it is (still) not illegal to do so (for example, fully automating a medical procedure without any human supervision) and/or 2) if it is better to do something, even if it is (still) not legally required (for example, implementing better working conditions for content moderators).

In this sense, the potential applicable safeguards will be addressed below.

3. State of the art of potential safeguards with immediate practical application

In an essay in the journal Nature (Yuste *et al*, 2017), Yuste and Goering, together with other medical professionals, ethics specialists, neuroscientists, and computer scientists, called for ethical guidelines to cover the increasing use of computer hardware and software to enhance or restore human capabilities. According to Yuste, Director of the Columbia Neurotechnology Center and member of the Data Science Institute, "we just want to ensure that this exciting new technology, which can revolutionize our lives, is used for the good of humanity."

As research yields surprising results, the authors identify four main threats: loss of individual privacy, identity, and autonomy, and the potential for increased social inequalities as corporations, governments, and hackers gain additional power to exploit and manipulate people.

In this sense, the potentials of neurotechnology reveal the need for innovative approaches to protect human rights. Unlike other technologies, such as AI or digital technologies, neurotechnology interacts directly with the human brain. According to the OECD, neurotechnology consists of devices and procedures used to access, monitor, investigate, evaluate, manipulate, and/or emulate the structure and function of individuals' neural systems. Thus, neurotechnology can directly map and alter human brain activity.

Neurotechnologies in immersive environments, for example, that has huge potential currently, refer to technologies that interact with the nervous system, including the brain and peripheral nerves. When applied to immersive environments, these technologies enhance the user's experience by creating more interactive and responsive environments. Here's how they apply:

1. Brain-Computer Interfaces (BCIs): These devices allow users to control virtual environments using their brain activity. For example, in a VR game, a BCI can translate a player's thoughts into actions within the game, making the experience more immersive and intuitive.

2. Virtual Reality (VR): VR creates fully immersive digital environments. When combined with neurotechnologies like BCIs, users can navigate and interact with these environments using their thoughts, enhancing the sense of presence and immersion.

3. Augmented Reality (AR): AR overlays digital information onto the real world. Neurotechnologies can be used to make these overlays more interactive, such as adjusting the information based on the user's focus or emotional state.

4. Mixed Reality (MR): MR blends real and virtual worlds, creating environments where physical and digital objects coexist. Neurotechnologies can enhance MR by providing more natural and intuitive ways to interact with both real and virtual elements.

5. Haptic Feedback: This technology provides tactile feedback to users, simulating the sense of touch. When combined with neurotechnologies, it can create more realistic and immersive experiences by responding to the user's neural signals.

The integration of neurotechnologies into immersive environments raises several ethical and regulatory concerns: 1) Privacy: The collection and use of neural data can raise significant privacy issues. Ensuring that this data is handled responsibly and securely is crucial; 2) Consent: Users must be fully informed about how their neural data will be used and must give explicit consente; 3) Security: Protecting neural data from unauthorized access and misuse is essential; 4) Psychological Impact: The long-term effects of using neurotechnologies in immersive environments on mental health need to be studied and regulated.

Based on this overview, Yuste calls on governments worldwide to create and protect new rights: neuro-rights.

The proposal of the Neurorights Foundation (2021) establishes five neuro-rights:

1. The Right to Personal Identity: Limits should be developed to prohibit technology from altering the concept of oneself. By connecting individuals' brains to computers, neurotechnology can blur the line between a person's consciousness and external technological inputs;

2. The Right to Free Will: People should be able to make and have control over their own decisions without the manipulation of unknown external neurotechnologies. This neuro-right is closely linked to personal identity. When we rely on external tools that interfere with our decisions, the human capacity to decide their future may be challenged;

3. The Right to Mental Privacy: All data obtained after measuring neuronal activity ("NeuroData") must remain private. Additionally, the sale, commercial transfer, and use of neural data should be strictly regulated;

4. The Right to Equal Access to Brain Enhancement: Guidelines should be established to regulate the development and application of neurotechnological mental enhancement at both international and national levels. These guidelines should be based on the principle of justice and ensure equal access for all citizens;

5. The Right to Protection Against Bias and Discrimination: Countermeasures to combat bias should be established, such as standard machine learning protocols. The algorithm design should include inputs from different user groups to address harmful biases.

In a seminal article, Ienca and Andorno had already stipulated the need to establish new safeguards in the era of neuroscience and neurotechnology (Ienca and Andorno, 2017), concluding that normative responses should not focus exclusively on civil liability issues but also on fundamental human rights issues.

Conclusions

In conclusion, this study underscores the importance of empirical research for assessing the technosocial impacts of neurotechnologies. By employing contextual risk assessments and existing Human Rights Due Diligence (HRDD) mechanisms, technology companies can proactively address potential risks and prioritize human rights before the widespread adoption of neurotechnological devices. This approach aligns with both national and international human rights frameworks, ensuring the protection of individuals and promoting social and technosocial responsibility.

Furthermore, the integration of these mechanisms demonstrates a commitment to ethical standards and corporate accountability in the rapidly evolving field of neurotechnology.

In this context, to promote the fulfillment of the United Nations Sustainable Development Goal (SDG) No. 16 (Peace, Justice, and Strong Institutions), with a focus on the rule of law, which is strongly interconnected with the right to development, this work aimed to enhance the discussion on the responsibility of technology companies. This responsibility extends not only to social but also technosocial aspects when employing and making available innovative technologies, whose risks and impacts on human rights are not yet well defined or addressed.

Given that the regulatory landscape is still emerging and requires cross-border approaches, corporate behavior becomes of utmost importance in guaranteeing human rights.

Despite the already established international historical framework for the protection of human rights, the deepening and potential impact of recent innovations in the field of neurotechnology, as mentioned above, guide neuroscientists and researchers, who have even driven these initiatives, to emphasize the need for additional safeguards, such as the observance of neuro-rights in human rights due diligence mechanisms, for example.

The society is at a turning point where companies launch products, and civil society and academia struggle to analyze the adverse technosocial impacts on health, well-being, opportunities, power asymmetries, and social inequalities affecting the population, especially vulnerable groups. Companies themselves should actually be responsible and accountable for the technosocial impacts of what they develop and commercialize.

This aims to contribute to a shift toward the construction of human and inclusive digital horizons, enabling cohesive and trustworthy social coexistence, where future generations can live and coexist with dignity. A collective project put into practice can never be labeled as utopian

References

B-TECH PROJECT. (2019). *OHCHR and business and human rights.* https://www.ohchr.org/en/business-and-human-rights/b-tech-project.

EUROPEAN UNION. (2021). *Ethics By Design and Ethics of Use Approaches for Artificial Intelligence.* https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf

Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions of the Royal Society* A: Mathematical, Physical and Engineering Sciences, v. 376, p. 2133. https://pubmed.ncbi.nlm.nih.gov/30322997/

Genser, j; Herrmann, s; Yuste, R. (2024). Protection Gaps in the Age of Neurotechnology. *NeuroRights Foundation.* https://neurorightsfoundation.org/reports

Hine, e., Rezende, i., Roberts, H. *et al.* Safety and Privacy in Immersive Extended Reality: An Analysis and Policy Recommendations. *Digital Society*. v. 3, n. 33. https://link.springer.com/article/10.1007/s44206-024-00114-1

ICO. – Information Commissioner's Office (2023). *Ico warns of "real danger" of discrimination in new Technologies that monitor the brain.* https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/06/ico-warns-of-real-danger-of-discrimination-in-new-technologies-that-monitor-the-brain#:~:text="We%20want%20to%20see%20everyone,the%20real%20danger%20of%20discrimination."&text=Discrimination%20in%20neurotechnology%20could%20occur,assumptions%20about%20people%20and%20communities.

ICO – Information Commissioner's Office. (2023). *Neurotechnologies: Key definitions*. https://ico.org.uk/about-the-ico/research-reports-impact-and-evaluation/research-and-reports/technology-and-innovation/ico-tech-futures-neurotechnology/neuro-technologies-key-definitions/

ICO – Information Commissioner's Office. (2023). UK. *ICO tech futures: neurotechnology.* 2023. https://ico.org.uk/about-the-ico/research-reports-impact-and-evaluation/research-and-reports/technology-and-innovation/ico-tech-futures-neurotechnology/

IEEE Neuroethics Framework - IEEE Brain. 2023. https://brain.ieee.org/publications/neuroethics-framework/addressing-the-ethical-legal-social-cultural-implications-of-neurotechnology/

Ienca, m.; Andorno, r. Ienca. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy. v.* 13, n. 1, p. 5-32 https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-017-0050-1

ISC - INTERNATIONAL Science Council (2023). *A framework for evaluating rapidly developing digital and related technologies:* AI, Large Language Models and beyond. Paris, France, International Science Council. https://council.science/publications/framework-digital-technologies/

Kalvaitytė, M. (2022). *Unregulated Negative Impacts of AI: Mixed Methods Analysis of Feedback Responses to the EU AI Act Proposal.* Public Policy Master Thesis. SciencesPo, School of Public Affairs. https://www.sciencespo.fr/public/sites/sciencespo.fr.public/files/KALVAITYTE.pdf

Lopes, Ana Maria D'Ávila Lopes et al (orgs) (2022). *Neurodireito, Neurotecnologia e Direitos Humanos*. Porto Alegre: Livraria do Advogado.

NEURORIGHTS Initiative (2021). *The Five Ethical NeuroRights*. https://neurorights-initiative.site.drupaldisttest.cc.columbia.edu/sites/default/ files/content/The%20Five%20Ethical%20NeuroRights%20updated%20pdf_0.pdf

OECD. (2019). *Recommendation of the Council on Responsible Innovation in Neurotechnology*. Doc. No. OECD/LEGAL/0457 https://www.oecd.org/science/recommendation-on-responsible-innovation-in-neurotechnology.htm

OECD. (2024). Framework for Anticipatory Governance of Emerging Technologies, *OECD Science, Technology and Industry Policy Papers*. n. 165, OECD Publishing, Paris. https://doi.org/10.1787/0248ead5-en

Salas, J. 2020. *Por que é preciso proibir que manipulem nosso cérebro antes que isso seja possível.* 13 fev. 2020. El País. https://brasil.elpais.com/ciencia/2020-02-13/por-que-e-preciso-proibir-que-manipulem-nosso-cerebro-antes-que-isso-seja-possivel.html

Perkis, A. (2020). QUALINET White Paper on Definitions of Immersive Media Experience (IMEX). https://www.academia.edu/64494985/QUALINET_White_Paper_on_Definitions_of_Immersive_Media_Experience_IMEx_?form=MG0AV3

UNESCO (2024). Outcome Draft Text First Version of A Draft Text of A Recommendation on The Ethics of Neurotechnology (SHS/BIO/AHEG-Neuro/2024/1) https://unesdoc.unesco.org/ark:/48223/pf0000389768

UNESCO (2023). *The risks and challenges of neurotechnologies for human rights*. https://unesdoc.unesco.org/ark:/48223/pf0000384185.

UE- UNIÃO EUROPEIA (2023). *EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition.* Communication from the Commission to the European Parliament, the Council, the european economic and social committee and the committee of the regions. {SWD(2023) 250} Strasbourg, 11.7.2023 COM(2023) 442/final. https://digital-strategy.ec.europa.eu/en/library/eu-initiative-virtual-worlds-head-start-next-technological-transition

Yuste, R., et al. (2017). Four ethical priorities for neurotechnologies and AI. *Nature.* v. 551, p. 159–163. https://www.nature.com/articles/551159a

Zuboff, S. (2020.) *A era do capitalismo de vigilância*: a luta por um futuro humano na nova fronteira do poder. George Schlessinger (Trad). Rio de Janeiro: Intrínseca.

— 9 —

The hardening of international soft law documents in the field of Neurolaw

ANA MARIA D'ÁVILA LOPES
Universidade de Fortaleza (Brazil)
anadavilalopes@unifor.br

Introduction

One of the biggest challenges that Law must face today is how to regulate advances in Neurotechnology, due to the speed with which they develop and the complexity of their impacts. Some international documents on the subject have already been approved, however, they are all soft law, that is, they are not mandatory, although they establish legal parameters to be followed by the States.

There are few states that have currently passed specific rules on advances in Neurotechnology, with Chile standing out as the first country in the world to reform its Constitution to introduce a rule to protect people from this type of advance. Nevertheless, this constitutional rule is a generic provision that still needs to be regulated by law, the drafting process of which began more than three years ago and has not yet been completed. Other countries, especially in Latin America, have begun the process of drafting rules, but they have not yet been a passed, at least as of the time of writing this paper.

This situation is concerning because several neurotechnologies are already being widely used, and not only in the field of Medicine. In the case of neurotechnologies for therapeutic purposes, the rules on medical treatments or procedures have been applied with some effectiveness to protect human beings. In the case of other types of neurotechnologies, their commercialization has been free, with practically no legal regulation, despite putting several human rights at risk.

In this context, this paper aims to propose the hardening of the provisions contained in international soft law documents on Neurotechnology, until hard law

documents are approved, as a way to protect human dignity. It is not, it must be recognized, a perfect solution, since international documents do not necessarily coincide on key points, such as the use or not of Neurotechnology for the purpose of improving human capabilities. However, in certain cases, it could be a valid alternative, as it would help to fill legislative gaps. To this end, bibliographical research was carried out in national and international doctrine, aiming to build the theoretical framework necessary for the understanding and analysis of the object of this study. Documentary research was also carried out on the main initiatives already undertaken by international organizations for the regulation of Neurotechnology.

The results of this research are summarized in this paper, which has been divided into three parts. The first part presents the rules necessary for the hardening of a soft law rule, as set out by former ECtHR judge Paulo Sérgio Pinto de Albuquerque in the Mursic v. Croatia case. The main international documents on Neurotechnology that have been approved or are in progress are then presented. Finally, based on the Emotiv case decided by the Chilean Supreme Court in 2023, the aim is to show the importance of hardening a soft law norm to fill regulatory gaps and consequently protect human rights.

1. The rules for the hardening of soft law: Mursic v. Croatia

On 20 October 2016, the European Court of Human Rights (ECHR) issued a judgment against the State of Croatia for violating Article 3 of the European Convention on Human Rights (ECHR), which prohibits torture and inhuman or degrading treatment or punishment (COUNCIL OF EUROPE, 1950). This is the case of Mursic v. Croatia (ECHR, 2016), in which the State was convicted of having held Kristijan Mursic in prison for 27 consecutive days in a cell with an individual space of less than 3 m2, in breach of the minimum European standards (Lopes, 2021).

Paulo Sérgio Pinto de Albuquerque, a judge at the ECtHR from 21 April 2011 to 31 March 2020, issued a partially dissenting opinion, considering that the Court erred in not considering the European Prison Rules (EPR), which establish a minimum of 4 m^2 (COUNCIL OF EUROPE, 2006). For Pinto Albuquerque, the EPR cannot be ignored by the ECtHR because they have the character of hardened soft law and, therefore, are binding.

In international law, hard law and soft law are categories used to indicate the binding or non-binding nature of a rule. Those that are mandatory are classified as hard law, while those that are not considered soft law. Recently, the ECtHR has recognized a third category, which would be the hardened soft law rules, the subject of this paper.

As Pinto de Albuquerque explains, soft law is a controversial concept. Its main criticisms denounce its imprecision, its use to disguise inflated international law or to merely proclaim political commitments to States that it does not wish to sanction or being a way of hiding the broad power of some countries over others.

Without ignoring these criticisms, Pinto de Albuquerque points out the advantages of soft law. Thus, when there is express legislation, it can be used to help understand the intention of legislators, facilitate the application of binding instruments,

clarify complex and technical points or adapt the law more easily to social changes. When there is no express legislation, soft law can act as a relevant practice of international law, being able to shape the *opinio juris* and state practice that make up customary international law. Despite its relevance, to date nothing has been done to clarify how the Council of Europe and, especially, the ECHR, address this issue.

However, Pinto de Albuquerque notes that it must be admitted that the ECHR cannot be interpreted in isolation but must be interpreted in harmony with international law and soft law, since there are no autonomous regimes in international law.

The ECHR, as a living instrument, says Pinto de Albuquerque, must be interpreted in the light of current conditions and the evolution of the rules of national and international law, as decided in the case of Tyrer v. the United Kingdom, in 1978.

He also clarifies that the evolutionary interpretation of the ECHR, from the beginning of the operation of the ECHR, "was closely linked to the need for a consensual reading of the text", based on the "common heritage of political traditions, ideals, freedom and the rule of law, to which the Preamble makes reference".

In this way, soft law constitutes, according to Pinto de Albuquerque, the realization of the European consensus and the common heritage of values that make up the internal legal framework of the great majority of the Member States, as decided in Marckx v. Belgium, in 1979. However, similarly to the interpretative technique used by the European Constitutional Court, the ECtHR modulates the effects of its decisions in order, for example, to avoid applying them retroactively, thus preserving legal certainty.

On the other hand, it explains that the interpretation of the ECHR must reflect the highest standard of international protection of human rights, as decided in Demir and Baykara v. Turkey, in 2008, meaning that the invocation of other treaties does not depend on their ratification by the parties, nor on the number of States-parties bound to them, demonstrating that hard law and soft law are intertwined in European law. Under this same logic, the ECtHR has also used rules from other Council of Europe bodies, which are not party to the ECHR.

The importance given to consensus in the Tyrer v. The United Kingdom Convention, from 1978, emanates, as Pinto de Albuquerque stated, from the understanding of the existence of "a deliberative, international democracy in which a majority or representative proportion of the Contracting Parties to the Convention is considered to speak in the name of all and is thus entitled to impose its will on other Parties", showing the difference between consensus and unanimity, that is, there may be consensus, even if not everyone agrees.

However, in the same §20, Pinto de Albuquerque recalls the criticisms raised against the ECHR in relation to its legitimacy to interpret the ECHR in an innovative way, "using soft law to circumvent the competent legislative bodies and to flout the principles of democracy, rule of law and subsidiarity", a criticism that, for him, can be considered obsolete, as it is inspired by the outdated sovereign leitmotiv of *in dubio pro mitius*.

For Pinto de Albuquerque, the Preamble of the ECHR determines that the objective of the Council of Europe is to create greater unity between the member States, based on common understanding and observance of human rights, with the principle

in dubio pro persona playing a fundamental role in this logic. In this sense, he clarifies that "evolutionary interpretation, European consensus and the hardening of soft law compose the three pillars of the European normative system", without this meaning dispensing with the consent of the States. The European legal order, he added, cannot continue to be reduced to an international pact of juxtaposed selfishness, nor to sovereignty as an absolute fact.

However, Pinto de Albuquerque recognizes that European human rights law is still tied to the formal theory of the sources of law. Soft law is not included as one of the classic sources of international law established in the Statute of the International Court of Justice (ECIJ), whose expansion requires a very demanding burden of proof. In any case, the list of sources established in article 38 of the ECIJ is not exhaustive, considering, especially, the changes suffered by international law after the Second World War, making it possible, therefore, to reformulate it.

Thus, based on the doctrine of the International Court of Justice, the ECHR has considered that, for a provision to become customary international law, it must meet the following conditions:

a) It must have a fundamentally regulatory character, so that it can serve as a basis for a Rule of Law.

b) It must be a practice consolidated by the States.

c) It must demonstrate "a belief that such practice is rendered obligatory by the existence of a Rule of Law requiring it (*opinio juris sive necessitatis*)".

On this, Pinto de Albuquerque clarifies that in the ECHR "the inherent deformalization of European human rights law is not a synonym of nihilist dissolution of legality". However, this "deformalization" of European human rights law should not be interpreted as a denial of legality, but as the adoption of a "cosmopolitan perspective of the universality of human rights and a dialogic understanding of the common heritage of values of European societies".

In this sense, he points out that the rule of recognition is no longer a voluntary, bilateral, top-down rule, centered on the State, but "democratic-type, individual-centred, broadly multi-lateral, purposefully consensual, bottom-up norm creation mechanism which involves European States and other European and non-European non-State actors. Distancing itself from an outdated *jus inter gentes"*. In any case, soft law is subject to the democratic control of the Council of Europe and its organs, which confirms its subsidiary nature and reinforces "the democratic nature of the process and the responsiveness of the international public policy-making system towards the European people".

This more complex understanding of the rule of recognition should not be interpreted as an increase in the degree of legal uncertainty, since, Pinto de Albuquerque warns, "The degree of legal certainty will rather depend on the substance of the legal discourse" than on aspects of form or procedure.

Between soft law and hard law, Pinto de Albuquerque adds, there is a gradual normativity derived from several factors. It is up to the ECHR to decide how these factors contribute to the hardening of soft law, which can refer both to the drafting procedure and to the application of the norms. These parameters are:

a) The prescriptive language adopted in a text must, in principle, be read as a normative text.

b) The degree of linguistic accuracy and the precision of the content of the text are indicative of its normative nature.

c) The existence of preparatory work, including an exhaustive analysis of causes and consequences of the political decisions taken, increases the normative density of the text.

d) The complexity of the deliberation process, including the voting pattern, is an indication of hardening.

e) The degree of publicity given to the normative text reveals the concern for its dissemination and consequent awareness and compliance.

f) The delegation to an independent third body of the interpretation and resolution of conflicts, as well as the provision of follow-up mechanisms are also indicative of hardening.

g) Subsequent practice, especially by the same authority, reinforces its normative claim.

The ECHR has understood that these factors are not just facts, or simple sociological reality, but that they transform soft law norms into sources of law, subject to the principle of *jura novit curia*.

For Pinto de Albuquerque, a hardened soft law norm produces the same legal effects as a hard law norm. Therefore, States are obliged to act in accordance with this hardened soft law norm, and their non-compliance may be considered a violation of international law. He also affirms that hardened soft law has a derogatory effect in relation to other laws, in addition to not being able to be "softened" again.

On this last point, Pinto de Albuquerque is very emphatic in stating that "Hard European human rights law may not be softened." Flexibility would imply going against the ECHR and the objective of the Council of Europe, which could lead to the suspension of the State, based on the provisions of art. 3 of the Statute of the Council.

Soft law is therefore a means by which European law can be developed, serving to enrich hard law or, if it does not exist, to fill this regulatory gap, which could be applied to all international law.

It is observed that the hardening of soft law presents itself as a valid alternative in cases of absence or deficiency of standards regulating a certain subject, as in the case of Neurotechnology, on which there is still no binding international document and few countries have approved any regulation on the subject, despite its impacts, and even violations, in the field of human rights.

2. International legal documents on Neurotechnology

In recent years, advances in Neurotechnology have been significantly enhanced thanks to its association with Artificial Intelligence, generating countless benefits for humanity, especially in the field of Medicine.

However, many of these advances have also been questioned not only for their collateral medical effects, but also for putting several aspects of human dignity at risk, which has attracted the attention of jurists concerned about possible violations of human rights.

This is how a new branch of Law, called Neurolaw, emerged, whose objective is to study the legal impacts derived from advances in Neurotechnology, in order to identify possible violations of human dignity and formulate juridical limits, to prevent scientific advances from overlapping the well-being of humanity (Lopes, 2023).

It was Taylor, Harp and Elliott (1991) who used, in the article "Neuropsychologists and neurolawyers" published in 1991, for the first time the terms "Neurolaw" and "neurolawyers". Since then, the subject has been growing rapidly, with proposals about new human rights, called neurorights. The most relevant proposals are from: a) Ienca and Adorno (2017), who defend the existence of four neurorights (cognitive liberty, mental privacy, mental integrity and psychological continuity); b) Yuste, Genser and Hermann (2021), who have formulated five neurorights (mental identity, mental agency, mental privacy, fair access to mental augmentation and protection from algorithmic bias).

As is to be expected, initiatives to understand and limit advances Neurotechnology are not only from doctrinal scope. International organizations also focus their attention on the subject.

The first of them was The Organization for Economic Co-operation and Development (OECD) which, on December 11, 2019, approved the "Recommendation of the Council on Responsible Innovation in Neurotechnology", with the objective of guiding governments and scientists on the ethical, legal and social aspects derived from these advances, but without intending to stop its development. In this sense, the following new principles were formulated:

a) Promote responsible innovation to face the challenges arising from advances in Neurotechnology in the health sector.

b) Prioritize safe evaluation of neurotechnology development and use.

c) Promote the inclusion of all people in advances resulting from health-oriented Neurotechnology.

d) Encourage scientific collaboration in Neurotechnology innovations between all countries, sectors and study areas.

e) Enable society's participation in deliberations about Neurotechnology.

f) Implement the capacity supervisory and advisory bodies to address new developments in Neurotechnology.

g) Protect personal brain data and other information obtained through Neurotechnology.

h) Promote a culture of administration and trust in Neurotechnology in public and private sectors.

i) Anticipate and monitor the potential unintentional and inappropriate use of Neurotechnology.

Months later, on December 15, 2020, the International Bioethics Committee (IBC) of the United Nations Educational, Scientific and Cultural Organization

(UNESCO) published the first draft report on ethical aspects about Neurotechnology, titled "Preliminary draft report of the IBC on ethical issues of neurotechnology".

In paragraph 1 of this document, it is highlighted that "brain activity is the basis of cognitive, affective or other brain states". It is also based on cerebral activity that is defined when someone is considered legally dead, in addition to the fact that it is from this activity that notions of human identity, freedom of thought, autonomy, privacy and human development are derived. The document also contains important definitions, such as Neurotechnology:

> 6. Neurotechnology is the field of devices and procedures used to access, monitor, investigate, assess, manipulate, and/or emulate the structure and function of the neural systems animals or human beings. These include: (i) technical and computational tools that measure and analyze chemical and electrical signals in the nervous system, be it the brain or nerves in the limbs. These may be used to identify the properties of nervous system activity, understand how the brain works, diagnose pathological conditions, or control external devices (neuro-prosthesis, 'brain machine interfaces'); and (ii) technical tools that interact with the nervous system to change its activity, for example to restore sensory input such as with cochlear implants to restore hearing or deep brain stimulation to stop tremor and treat other pathological conditions. They are meant to either record signals from the brain and 'translate' them into technical control commands, or to manipulate brain activity by applying electrical or optical stimuli (UNESCO, 2020, par. 6).

In paragraph 13 it is stated that all these technologies should not only improve individual well-being but also produce positive effects for the community and society.

The importance of human dignity is stated in paragraph 37: "Having dignity means having a value to be recognized, respected and protected from arbitrary alteration, modification, manipulation, which violates it and causes harm to the subject (who becomes an object)" (UNESCO, 2020, par. 37).

In the following paragraphs are exposed the different ways in which human rights can be affected by advances in Neurotechnology, concluding that States and private entities must harmonize technological development with respect for human rights.

On May 19, 2021, the UN Human Rights Council approved the report of its Advisory Committee, entitled "Possible Impacts, opportunities and challenges of new and emerging digital technologies with regard to the promotion and protection of human rights", in which seven risks for human rights are indicated:

a) The loss of privacy and the need to protect personal data because of datafication.

b) Poor Cybersecurity and pervasive and deeply integrated digital systems.

c) Loss of quality and authenticity of information

d) Increased radicalization, segregation and discrimination

e) Disempowerment and inequality

f) Massive surveillance and excessive regulation of the Internet.

g) Diffusion of cyberviolence

Another relevant international initiative is the "Declaration of the Inter-American Juridical Committee on Neuroscience, Neurotechnology and Human Rights: new legal challenges for the Americas", prepared and approved by the Interamerican Juridical Committee of the Organization of American States (OAS) on August 11, 2021,

in which are listed six concerns about the development of Neuroscience and Neurotechnology, and its impacts on human rights:

> 1) Conditioning of personality and loss of autonomy due to indiscriminate and unregulated use of applications or technological devices, which can manipulate emotions, feelings and decisions, and also decode neuronal information, opening the possibility of affecting the privacy, dignity and identity of human beings.
>
> 2) Illegitimate interventions in matters of health, physical and mental integrity shows the importance to respect the right to informed consent and medical confidentiality. However, in the field of Neurotechnology could be difficult to recognize what constitute an illness or a cognitive improvement.
>
> 3) Obtaining neuronal data using Neurotechnology can put at risk the right to mental privacy. In this sense, human beings must be protected against malicious conduct by those who access brain activity data to penetrate their mind, condition it or take advantage of knowledge.
>
> 4) Equal access and non-discrimination constitute fundamental principles of the Inter-American System, which are also applicable to the use of Neurotechnology. Therefore, it is necessary to consider the adverse and disproportionate impacts of these advances on certain traditionally disadvantaged human groups, which can deepen existing inequalities.
>
> 5) The rights of freedom of expression and access to public information cannot be subject to limitations beyond those permitted by Law. Neurotechnology, just like Artificial Intelligence tools, must be intelligible to developers, medical professionals, patients, users and regulators. The proliferation of fake news and speculation constitutes a serious barrier to informed public debate.
>
> 6) Recommendations for the States, the private sector, the academy and the scientific world. In the case of States, they are called to anticipate and adopt the necessary measures to avoid negative impacts. It must also promote public policy actions that generate awareness and education, as well as establishing that the use of technologies based on the brain-computer interface must only serve legitimate purposes, prohibiting their use for the purposes of social control, mass surveillance of citizens or the like. On the other hand, the private sector must develop practices compatible with human rights, generate commitments, standards and good practices and incorporate the UN Regulatory Principles on Business and Human Rights (Ruggie Principles). In the case of academia, it must incorporate activities that allow understanding the impacts of Neurotechnology, as well as its scientific, ethical and social reach, in addition to promoting analysis and legal reflection on this matter.

In the end, the document calls on the scientific world to carry out a rigorous evaluation of the risks and benefits of advances in Neurotechnology, in addition to the need to promote instances of scientific dissemination to large audiences.

Following this line of concerns, on April 12, 2022, the UN Human Rights Council published a report prepared newly by its Advisory Committee, which, in Annex III, which deals with research proposals, highlights the evaluation of the impacts of Neurotechnology on human rights ("Assessing the human rights impact of neuro-

technology: toward the awareness of neurorights"), discussing the need for a new of human rights.

> D. A new challenge for human rights
>
> From a human rights perspective, it has been argued that existing instruments are ill-equipped to face these new challenges.
>
> Experts suggest that a new set of rights should be recognized with a view to introducing "normative specifications related to the protection of the person's cerebral and mental domain", which includes individual mental integrity and identity. Here the question is whether such protection should be introduced as a new human rights norms or rather as standards of application or interpretation of existing rights, while reinforcing the applications of the principles on business and human rights and other specific initiatives in parallel.
>
> This assessment requires a careful and balanced analysis of the new norms that are being proposed as rights. It is true that specific standards may be needed to ensure the protection against interferences and misuses of certain mental aspects such as cognitive freedom, mental privacy, mental integrity and psychological continuity. The equal access to neurotechnology for medical purposes has also been promoted, together with individual's access to justice and adequate accountability mechanisms. There are, however, other much more disputable interpretations, such as the claim that a right to fair access to "mental augmentation" should be recognized. (UN, 2022b, Annex III – p.3-4).

This concern about the impacts of Neurotechnology on human rights led to the approval, on October 6, of Resolution A/HRC/RES/51/3 on "Neurotechnology and Human Rights" during the 51st Session of the Human Rights Council, held from September 12 to October 7 this year. (UN, 2022a). In this document, the Council requests that the Advisory Committee prepare an accessible study on

> [....] the impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights, including recommendations on how human rights opportunities, challenges and gaps arising from neurotechnology could be addressed by the Huma Rights Council and its special procedures and subsidiary bodies in a coherent, holistic, inclusive and action-oriented manner, and to present the study to the Council at its fifty-seventh session (UN, 2022a, par. 1).

The elaboration of this study, which must include an easy-to-understand text, must involve a broad participation of global society, including representatives of States, international organizations, human rights institutions, civil society, the private sector and academic, technological and scientific institutions.

On March 9, 2023, the CIJ of OAS approved a second document on the subject, entitled "Inter-American Declaration of Principles regarding Neuroscience, Neurotechnologies, and Human", which contains concrete guidelines, considering that these advances can "affect the essence of [their] personality and identity, resulting in important ethical-legal challenges for the guarantee of enshrined Human Rights", evidencing a clear concern on the subject. In this document, 10 principles are listed, elaborated based on current standards of human rights adapted to the field of Neurotechnology:

a) Principle 1: Identity, autonomy and privacy of neural activity. It must be preserved and ensured that each person can decide on their own individual identity, as well as on their self-determination and freedom of thought.

b) Principle 2: Protection of human rights in the design of Neurotechnology. States should promote a human rights-based approach in the research, development, implementation, commercialization, evaluation and use of neurotechnologies.

c) Principle 3: Neural data are sensitive personal data. States should promote measures to ensure the control, security, confidentiality and integrity of neural data, including limits on the use of decoding techniques that allow a person to be identified or made identifiable, especially with databases or sets of information that are shared with third parties.

d) Principle 4: Express and informed consent regarding neural data. Free, informed, express, specific, unambiguous and flawless consent must be guaranteed when it comes to accessing or processing neural activity. Consent must be revocable at any time. Children and adolescents, as well as people with disabilities, the elderly and people deprived of liberty must be given special protection

e) Principle 5: Equality, non-discrimination and equal access to Neurotechnology. States, based on the principle of equality and non-discrimination, must ensure equitable access to neurotechnologies, respecting customs and traditions, and avoiding exacerbating inequality and discrimination, especially against the most vulnerable groups.

f) Principle 6: Exclusive therapeutic application in relation to the improvement of cognitive abilities. States must exercise caution in regulating the use of Neurotechnology to increase individuals' cognitive abilities and must establish clear limits.

g) Principle 7: Neurocognitive integrity. It is essential to prevent the use of Neurotechnology for malicious purposes that may harm brain activity or impair the exercise of human rights. Every person has the right not to suffer violations, alterations, manipulations and/or modifications of their neurocognitive integrity and privacy that put personal integrity at risk or harm it.

h) Principle 8: Transparent governance of neurotechnology. States must ensure transparency of all those involved in the study, development and application of Neurotechnology, as well as its impact on human rights, in addition to accountability for the use of neural data.

i) Principle 9: Supervision and control of Neurotechnology, States must designate a competent, technically specialized and financially autonomous national authority to supervise neurotechnology. They must also pay special attention to the rights of children and adolescents, people with disabilities and people deprived of their liberty.

j) Principle 10: Access to effective protection and remedies in the face of the development of Neurotechnology. States must guarantee access to remedies, including judicial mechanism and compensation for human rights violations.

More recently, UNESCO has also developed a document on the subject. Therefore, at the 42nd session of the UNESCO General Conference held in April 2024, an Ad Hoc Expert Group (AHEC) was established, through Resolution 29 (42C/Resolution 29), to prepare a preliminary text of the Recommendation on Ethics in Neurotechnology.

The first meeting of the AHEG, composed of 24 specialists from different countries and areas of knowledge (Medicine, Law, Ethics, Neurotechnology, etc.) was held between 22 and 24 April 2024, at the UNESCO headquarters in Paris, where the

first version of the project was prepared, which was submitted to national and regional consultations in the months of June and July 2024. The second AHEG meeting was held in August and the document is expected to be approved in November 2025.

The first version of the Recommendation on Ethics in Neurotechnology, of UNESCO dated May 9, 2024 (UNESCO, 2024), is divided into 7 parts: a) Preamble; b) Definitions and Field of Application; c) Our approach; d) Values, Principles and Rights; e) Areas of Political Actions; f) Implementation; g) Promotion of this Recommendation and Final Provisions. It is an extensive document, of more than 30 pages, with some relevant points, such as the definition of Neurotechnology in paragraph 2.

> 2. Neurotechnology refers to devices and procedures used to understand and/or influence, access, monitor, assess, emulate or modulate the structure and function of the nervous systems of human beings and other animals. While neurotechnology has great potential to significantly enhance human experiences and functioning, it also raises important ethical, legal and societal concerns related to security, mental privacy, cognitive liberty, agency and human autonomy (UNESCO, 2024, par. 2).

It is observed that, in this definition, animals are also included, although in paragraph 14 it is made clear that the Recommendation is only directed to human beings. On the other hand, the concern about minorities and groups in situations of vulnerability, such as indigenous communities, children and people with disabilities is present in several parts of the text.

A controversial aspect of this document that should be highlighted is found in paragraphs 168 to 172, in which there is a reference to neurotechnology for enhancement, which, although they specify the need to avoid discrimination and the increase in inequalities, is accepted, which does not happen, for example, in documents prepared by the OAS, in which the therapeutic use of Neurotechnology is expressly established.

In these initiatives of the OECD, OAS, UN and UNESCO it is possible to identify clear signs of recognition of the urgency of regulating advances in Neurotechnology (Lopes, 2024), however, all these documents are considered soft law and, therefore, are not imperative, which constitutes a serious deficiency in the face of the accelerated development of Neurotechnology and concrete experiences of violation of human rights by these advances, as explained in the following section.

3. The Emotiv case and the need to hardening soft law in the field of Neurotechnology

On August 9, 2023, the Supreme Court of Chile suspended the commercialization of the Insight neurotechnology device, of the North American company Emotiv, considering that it did not adequately protect the brain information and the physical and mental integrity of its users (Chile, 2023). This judicial decision is, without a doubt, a fact of unquestionable relevance for the protection of neurorights, insofar as it is the first time in the world that a Supreme Court has ruled on this matter.

However, it is not the first time that Chile has stood out on the world stage for innovating in the protection of neurorights. On October 25, 2021, it became the first country in the world to reform its Constitution to incorporate a rule for protecting neurorights. The reform of the Constitution of 1980 was carried out through the Law n. 21,383 (Chile, 2021) that incorporated, in Article 19, a rule to protect brain activity

from scientific and technological development, establishing that this type of activity must be at the service of people, as well as respect their life and physical and mental integrity. It also provided that a law will regulate the requirements, conditions and restrictions for the use of these advances, especially safeguarding brain activity, as well as the information derived from it.

Months earlier, on October 7, 2020, a bill regulating advances in Neurotechnology had already been proposed in the Chilean Senate (Law Bulletin n. 13,828-19), which approved it and sent it to the Chamber of Deputies on December 7, 2021, where it is still awaiting approval (Chile, 2020).

These two Chilean initiatives not only reveal the concern of its authorities to protect neurorights but are also a reflection of the intense activity of the academic community in that country, which has been debating for some time now the need to regulate advances in Neurotechnology, considering its capacity to put human rights at risk.

The Insight device is a clear example of this, insofar as it is a tool that not only has the capacity to directly access the user's brain data - even without the user being aware of it - but also to analyze and interpret this data, being capable of influencing and altering human behavior and, again, without the user being necessarily conscious of this, for commercial, political, educational, therapeutic purposes, etc. with important benefits, but also with unquestionable risks to integrity, freedom, identity, privacy and equality (Lopes, 2023).

There are also improved neurotechnologies being developed with the capacity to change human nature itself, as is the case of implantable chips in the nervous system, which can even have a eugenic character, such as the recent implantation of a chip in the brain of a quadriplegic man by the company Neuralink (Paredes; Lopes, 2024).

These technological innovations require special attention from the State as all their effects are now completely unknown, as the Chilean Supreme Court warned in the sentence of the Emotiv case:

> [...] before the development of new technologies that increasingly involve more aspects of the human person, aspects that were unthinkable a few years ago that could be understood, special attention and care must be given to their revision by the State, with the aim of preventing and anticipating their possible effects, in addition to directly protect human integrity in its totality, consideration that includes its privacy and confidentiality and the proper rights of psychic integrity and the subject of scientific experimentation (Chile, 2023, p. 8, our translation).

This is, without a doubt, an emblematic decision, however, it should be noted that it is a sentence of only a few pages (barely 8) that, despite emphasizing the need to protect human beings against two advances in new technologies, takes a cautious position, since it transfers to the National Customs Service and the Public Health Institute the analysis to determine whether the Insight device complies with the regulations in force in Chile.

> [...] declares that the appeal for protection is accepted for the sole purpose of having the Public Health Institute and the customs authority evaluate the background in the use of their powers, making the corresponding legal arrangements, so that the marketing and use of the Insight device and the handling of data obtained from it strictly conform to the regulations applicable in the case and outlined in this judgment Chile, 2023, p. 8, our translation).

On August 9, 2023, when the Supreme Court of Chile ruled, several international documents on the use of neurotechnologies had already been approved (as explained in the previous topic of this paper) and, although the Chilean Court cited

in its decision some general international human rights documents or specific ones, such as those on the human genome, it omitted any reference to documents in the area of Neurotechnology, which could have helped to better support the ruling and could even have helped to clearly demonstrate the potential of these devices to put neurorights, such as privacy and integrity, at risk, without the need to transfer this responsibility to the customs or health authorities.

The Chilean Supreme Court missed the opportunity to develop deeper into the discussion on the ethical and legal limits of Neurotechnology, using documents already approved by the OECD, OAS, UN and UNESCO, which would also have contributed to considering the hardening of the provisions of these documents, as explained by former ECtHR judge Paulo Sérgio Pinto de Albuquerque in the Mursic v. Croatia case, as discussed in the first topic of this paper.

However, it is important to be cautious about which provisions in these documents should be hardening, given that there are still many controversies on essential topics in the field of Neurotechnology, such as its use for enhancement purposes. It should be noted that the 2023 OAS Declaration prohibits it, but the UNESCO document that is being prepared allows it, showing the need for greater academic dialogue, which can also expand throughout global society, so that everyone's opinions can be heard.

Conclusion

In recent years, several international documents on Neurotechnology have been approved. However, all of them are soft law, that is, they are not binding. This is not necessarily a negative situation, as it may simply reveal that we are dealing with a new and complex issue that requires greater debate and caution to avoid creating rules that may be insufficient or deficient, especially considering the speed with which advances in Neurotechnology are developing.

The truth is that, until binding rules are approved, several neurotechnologies are being freely commercialized, putting several human rights at risk.

In view of this situation, this paper sought to propose the hardening of the provisions present in the current international documents on Neurotechnology, as a way of filling the legislative gaps in national legal systems.

It is certainly not a perfect solution, because there is no uniformity in these documents on some crucial points on Neurotechnology, such as its use for enhancement purposes, which, in addition to its possible eugenic character, can irremediably aggravate inequalities between human beings. Therefore, the rules on the hardening of soft law rules, proposed by former ECtHR judge Paulo Sérgio Pinto de Albuquerque in the Mursic v. Croatia case, must be followed carefully, without forgetting that science must be at the service of all humanity and not just some human beings.

References

Chile (2021). *Ley núm. 21,383.* Protección de los neuroderechos y la integridad mental, y el desarrollo de la investigación y las neurotecnologías. Promulgada el 25 de octubre de 2021. https://www.bcn.cl/leychile/navegar?idNorma=1166983

Chile. (2020). *Boletín Nº 13.828-19*. Projeto de Ley sobre la protección de los neuroderechos y la integridad mental, y el desarrollo de la investigación y las neurotecnologías, propuesto el 7 de octubre de 2020. https://www.camara.cl/legislacion/ProyectosDeLey/tramitacion.aspx?prmID=14385&prmBOLETIN=13828-19

Chile. Corte Suprema (2023). *Girardi/Emotiv INC: 09-08-2023* ((Civil) Apelación Protección), Rol N° 105065-2023. https://juris.pjud.cl/busqueda/pagina_detalle_sentencia?k=L2FadjlzYUdmbGx1UWNKWmV6ZjZ3dz09

Council of Europe. *European Penitentiary Rules*. January 11, 2006. https://rm.coe.int/16804cc2f1

Council of Europe. *European Convention on Human Rights,* November 4th, 1950. https://www.echr.coe.int/documents/d/echr/convention_ENG

ECHR - European Court of Human Rights. *Mursic v. Croacia,* Application nº 7334/13, October 20, 2016. Partly dissenting opinion of Judge Pinto de Albuquerque, §1º. https://hudoc.echr.coe.int/eng#{"fulltext":["mursic"],"documentcollectionid2":["GRANDCHAMBER","CHAMBER"],"itemid":["001-167483"]}

CIJ – Inter-American Juridical Committee of the Organization of American States. *Inter-American Declaration of Principles regarding Neuroscience, Neurotechnologies, and Human Rights*, March 9, 2023. CJI/RES. 281 (CII-O/23). https://www.oas.org/en/sla/iajc/docs/CJI-RES_281_CII-O-23_corr1_ENG.pdf

CIJ – Inter-American Juridical Committee of the Organization of American States. Declaration of the Interamerican Juridical Committee on Neuroscience, Neurotechnology and Human Rights: new legal challenges for the Americas, August 11, 2021, CJI/DEC. 01 (XCIX-O/21). https://www.oas.org/en/sla/iajc/docs/CJI-DEC_01-XCIX-O-21_ENG.pdf

Taylor, J. Sherrod; Harp, J. Anderson; Elliot, Tyron. (1991). Neuropsychologists and neurolawyers. *Neuropsychology*, v. 5, n. 4, p. 293–305. https://doi.apa.org/record/1992-25630-001?doi=1

Ienca, Marcello; Andorno, Roberto. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life sciences, society and policy*, v. 13, n. 5, p. 1-27. https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-017-0050-1

Lopes, Ana Maria D'Ávila. (2024). Iniciativas internacionais de proteção aos neurodireitos humanos: transumanismo ou bioconservadorismo? *Quaestio Iuris*, Rio de Janeiro, v. 17, n. 1, p. 444-465. https://www.e-publicacoes.uerj.br/quaestioiuris/article/view/75061

Lopes, Ana Maria D'Ávila. (2022). Neurotecnologia: quando a tecnologia ameaça a dignidade humana. In: Lopes, Ana Maria D'Ávila et al. (org.); *Neurodireito, Neurotecnologia e Direitos Humano*s. Porto Alegre: Livraria do Advogado, p. 19-29. https://www.researchgate.net/publication/368128401_Neurodireito_Neurotecnologia_e_Direitos_Humanos

Lopes, Ana Maria D'Ávila (2021). Mursic vs. Croácia: contribuições do juiz Paulo Sérgio Pinto de Albuquerque para o endurecimento da soft law prisional. In: RAMOS, André de Carvalho. (Org.). *Corte Europeia dos Direitos Humanos*: Comentários dos votos do juiz Paulo Sérgio Pinto de Albuquerque. 1ed.São Paulo: Tirant lo Blanch, p. 246-264.

Paredes Paredes, Felipe Ignacio; Lopes, Ana Maria D'Ávila. (2024). Las limitaciones de los actuales debates jurídico-filosóficos sobre los avances neurotecnológicos: el caso neuralink. In: Pintarelli, Camila; Piovesan, Flávia; Abboud, Georges. *Constitucionalismo digital e direitos humanos*. Desafios da internet, inteligência artificial e neurotecnologia. São Paulo. Thomson Reuters, p. 371-388.

UN – United Nations. Human Rights Council Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development Fifty-first session 12 September – 7 October 2022, A/HRC/RES/51/3. *Resolution adopted by the Human Rights Council on 6 October 2022b*. Neurotechnology and human rights, October 13, 2022a. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/525/01/PDF/G2252501.pdf?OpenElement

UN – United Nations. Human Rights Council. Advisory Committee. Twenty-eighth session. 21-25 February 2022. *Assessing the human rights impact of neurotechnology*: towards the recognition of neurorights. A/HRC/AC/27/2, April 12 2022b. https://www.ohchr.org/en/hr-bodies/hrc/advisory-committee/session28/index

UN – United Nations. Human Rights Council Advisory Committee. Possible impacts, opportunities and challenges of new and emerging digital technologies with regard to the promotion and protection of human rights; A/HRC/47/52, May 19, 2021. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/110/37/PDF/G2111037.pdf?OpenElement

UNESCO - Organização das Nações Unidas para a Educação, a Ciência e a Cultura. *First Draft Text of UNESCO Recommendation on Ethics of Neurotechnology*. SHS/BIO/AHEG-Neuro/2024/, Paris, 9 May 2024. https://unesdoc.unesco.org/ark:/48223/pf0000391074

UNESCO - Organização das Nações Unidas para a Educação, a Ciência e a Cultura. *Preliminary draft report of the IBC on ethical issues of neurotechnology*, Paris 15 december 2020. SHS/BIO/IBC-Ext/2021/3 https://unesdoc.unesco.org/ark:/48223/pf0000375237

Yuste, Rafael; Genser, Jared; Herrmann, Stephanie. (2021). It's Time for Neuro-Rights: new human rights for the age of neurotechnology. *Horizons*, n. 18, p. 154-164. https://www.cirsd.org/en/horizons/horizons-winter-2021-issue-no-18/its-time-for-neuro--rights

— 10 —

Strategic human rights litigation against Artificial Intelligence companies[1]

SEBASTIÁN SMART
Anglia Ruskin University
Sebastian.Smart@aru.ac.uk

ESTEBAN OYARZÚN GÓMEZ
CEDIDH
estebanoyarzungomez@gmail.com

Introduction

As regulations on neurotechnologies expand globally, we expect to see a considerable increase of lawsuits related to these technologies. Currently, countries such as Chile, with a pioneering constitutional reform, Brazil, Spain and California have implemented or are developing regulations to regulate neurorights (Ligthart et al., 2023). In turn, the NeuroRights Foundation is promoting an international campaign to redraw the traditional concept of privacy, seeking to include not only undue interference, as it has been interpreted historically, but also new forms of intervention, such as those caused by neuroscience technologies (Yuste, 2023). This new framework aims to protect fundamental aspects of individuals' mental and cognitive integrity, which may be threatened by the advancement of neurotechnologies and Artificial Intelligence (AI).[2]

[1] This chapter was funded by the Fondecyt Initiation Project, "Strategic Litigation in the Field of Human Rights and Business" (No 11220195).

[2] For the purpose of this chapter we use the OECD definition of AI. AI system is a machine-based system that for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as makes predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Despite these regulatory advances, case law in this field is still limited. To date, the only case with an enforceable judgment at the international level related to neurotechnologies is the Girardi case, which we will address in this chapter. In this analysis, we will review what worked and what did not work in terms of strategic human rights litigation against companies developing technologies of this type. Although this case is an important precedent, it also presents a number of limitations and challenges. Based on this analysis, we seek to expand the debate to a broader realm: human rights litigation against AI companies. AI is not only key in the development of neurotechnologies, but it is also fundamental to understanding how these technologies can affect the neurocognitive rights of individuals.

In this chapter, we propose a theoretical approach based on a literature review of strategic litigation in human rights (from different fields), adapted to the field of neurotechnologies and AI. We will begin by exploring the traditional concepts of strategic litigation, using frameworks such as political disadvantage theory, which has been fundamental in human rights cases in contexts of vulnerability. In addition, we will review approaches based on legal resources and political opportunities, which have allowed certain actors to mobilize judicially in contexts of inequality. These theories will be the starting point for, in a second part of the chapter, to analyze how these conceptual frameworks can be applied or adapted to the context of litigation in AI and neurotechnologies.

Unlike previous studies that have focused on the use of AI in judicial systems, our goal is different. We will not focus on how these technologies transform the administration of justice, but on how theories of strategic litigation can be applied in cases against technology companies that use AI. The nature of AI poses new challenges for human rights, and companies developing these technologies are beginning to face litigation questioning their impact on privacy, cognitive manipulation, and algorithmic discrimination, as we will observe in the third part of this chapter.

The analysis of these cases will allow us to evaluate what has worked and what has not in terms of legal strategies. Mental privacy, for example, is an emerging right that has not yet been fully protected by existing laws, and strategies for litigating this right may require innovative approaches. In this sense, strategic litigation theories that have been successful in areas such as environmental law or indigenous rights can offer useful tools for lawyers and human rights defenders facing cases in the field of AI and neurotechnologies.

The impact of this work on academia and legal practice can be significant. By adapting existing theories of strategic litigation to an emerging field such as neurotechnologies, we provide a roadmap for future litigation and legal strategies. This can open new possibilities for the protection of human rights in a context where technologies advance faster than regulations. Likewise, by highlighting the use of AI in the violation of neurocognitive rights, this work can influence the development of more comprehensive regulatory frameworks that protect individuals against the risks of these technologies.

This chapter not only seeks to contribute to the theory, but also to provide practical tools for those who wish to undertake strategic litigation in this field. While human rights litigation has always been complex, the advent of new technologies such as AI and neurotechnologies requires a reimagining of traditional strategies. By identifying the challenges and opportunities in neurorights litigation, we hope to lay

the groundwork for more effective litigation with greater impact on the protection of human rights in the future.

1. The case of Girardi v. EMOTIV Inc.

The case of Girardi v. EMOTIV Inc. marked a before and after in neurotechnology litigation, as it was the first internationally enforceable ruling that addressed the protection of brain data within the framework of human rights (Smart and Oyarzun, 2024). This case originated in Chile, where a group of citizens, concerned about the improper handling of information captured by neurotechnology devices marketed by EMOTIV Inc., filed an appeal for protection before the Court of Appeals of Santiago. The device in question, called Insight, allowed the recording and storage of users' brain activity, generating serious doubts about privacy guarantees and the potential misuse of that data. The appellants argued that EMOTIV did not offer the necessary measures to guarantee the security of brain information, thus exposing users to the violation of their fundamental rights, such as dignity and mental integrity (Corte Suprema, 2013).

In the first stage of the trial, EMOTIV's defense focused on demonstrating that users had given their informed consent when using the device, and that the data collected was properly protected through encryption processes. However, the Supreme Court accepted the appeal, establishing that the company should review and adjust its data handling policies to align with the standards established by Chilean law. Through this ruling, the Court implicitly recognized the need to protect mental privacy, setting a precedent for how neurotechnologies should operate within the limits of human rights protection in Chile.

One of the most notable features of the Girardi case was the strategic impact it had, both nationally and internationally. Although it was an individual protection resource, the case was used as a strategic litigation to promote the regulation of neurotechnologies and neurorights. Organizations such as the NeuroRights Foundation took this ruling as an example of how the courts can play an important role in protecting fundamental rights in the face of technological advances. This generated greater discussion about regulatory development in other jurisdictions, particularly in Latin America, where countries such as Brazil and Argentina have begun to debate the regulation of these technologies (Smart and Oyarzun, 2024).

However, despite the legal success of the ruling, the case also revealed a number of important limitations. Although the ruling forced EMOTIV to review its practices, it did not comprehensively address the normative content of neurorights, nor did it establish clear criteria for the protection of brain data in future contexts. The ruling, while acknowledging the vulnerability of fundamental rights to the misuse of neurotechnologies, was limited to a general policy review order and did not impose specific sanctions or create a comprehensive regulatory framework.

Another criticism that can be made of the Girardi case is the lack of a solid response from the Chilean State regarding the effective implementation of the measures ordered by the Court. Instead of creating a specialized institutional framework to regulate and oversee the use of neurotechnologies, the Supreme Court delegated the task of evaluating EMOTIV's practices to institutions that lacked expertise in these issues, such as the Institute of Public Health (Corte Suprema, 2013). This re-

vealed a lack of institutional capacities to manage the implications of neurotechnologies and showed that digital governance in Chile was not yet prepared to face this type of challenge.

On a practical level, the ruling did not produce significant changes in Chilean legislation either. Although the appeal was accepted and generated an important legal precedent, the effects in terms of public policies and protection of rights in the long term have been limited. There have been no structural reforms, or the creation of new supervisory bodies specialized in neurotechnologies, which reduces the impact of the Girardi case at the institutional level(Smart and Oyarzun, 2024). In fact, the implementation of this ruling has not generated a solid framework that allows for the effective protection of neurorights, which weakens its ability to influence the future regulation of this technology in Chile and the region.

The limited impact of the case in practice demonstrates that while strategic litigation can be a powerful tool for driving human rights change, its success depends largely on the state's ability to implement judicial decisions. In this sense, the Girardi case illustrates the gap between judicial decisions and their effective implementation. In addition, it highlights the need for strategic litigation to be accompanied by a comprehensive regulatory framework and institutions capable of enforcing decisions appropriately (Smart and Oyarzun, 2024).

With the Girardi case in mind, it is crucial to discuss in the next section the theories that explain strategic litigation in human rights, as this case represents both the achievements and limitations of this type of litigation in the field of neurotechnologies. By examining theoretical frameworks that have been used in other human rights domains, we will be able to identify what elements work in contexts of technological vulnerability, and what adaptations need to be made to address the specific challenges presented by AI and neurotechnology companies.

2. Explaining strategic legal mobilization

Having analyzed the limitations of the Girardi case before the Court of Appeals of Santiago and the Supreme Court of Chile, we now want to present a theoretical framework to understand strategic legal mobilization. Much of the existing research on legal mobilization, by both individuals and groups, attempts to explain the shift to the courts and the conditions under which legal mobilization is likely to succeed, with many definitions of "success" deployed. Political scientists and sociologists do not shy away from speculating on the general causes and consequences of legal mobilization. This is the case, even though much of the research is based on case studies, i.e. with a focus on particular legal cases or areas of policy issues. Although the discussion here separates the literature according to the type of theoretical explanation proposed, this categorization is not strict. At present, most research transcends some or all the factors highlighted by the various conceptual approaches that we will develop in this section.

2.1. Political disadvantage theory

Political disadvantage theory holds that marginalized groups who lack access to traditional political mechanisms turn to the courts to seek protection and advance

their goals. Initial studies, such as those of Cortner (1968) and Vose (1959), who analyzed legal mobilization during the civil rights movement in the United States, showed that groups without influence in executive or legislative bodies found in the courts an avenue to assert their rights. This approach was widely accepted until the 1980s, when studies such as Epstein's began to show that even politically privileged groups could use the courts strategically, leading to criticism of the widespread applicability of this theory (Epstein, 1985).

At the national level, marginalized groups often rely on litigation and alternative dispute resolution (ADR) mechanisms to address systemic inequalities. For example, in Canada, First Nations groups pursued justice for discriminatory child welfare policies through the Canadian Human Rights Tribunal (Hernandez and Levesque, 2023). Similarly, Dalits in India, a historically marginalized group, have turned to courts through NGO-supported litigation to combat caste-based discrimination, such as the inhumane practice of "manual scavenging" (Kahle and Hammerslev, 2023). In some cases, ADR mechanisms such as mediation and conciliation have provided an alternative pathway to justice for marginalized groups unable to access formal court systems (McGregor et al., 2019). However, ADR is not without challenges. Critics argue that when ADR becomes mandatory, it can exacerbate political disadvantage by limiting marginalized groups' access to full judicial remedies. For example, McGregor (2012) warns that ADR processes, while efficient, may prioritize expedience over the structural changes needed to address systemic inequality, leaving vulnerable groups without the tools to challenge deeply rooted injustices.

Regional human rights systems also provide vital avenues for politically disadvantaged groups. The European Court of Human Rights (ECtHR) offers prominent examples of how groups excluded from political mechanisms use courts to assert their rights. In *Kokkinakis v. Greece* (1993), Jehovah's Witnesses successfully challenged bans on religious proselytism, while in *Dudgeon v. United Kingdom* (1981), the ECtHR addressed the criminalization of homosexuality in Northern Ireland (Anagnostou, 2010). Similarly, in *D.H. et al. v. Czech Republic* (2007), Roma children's school segregation was challenged, underscoring the role of courts in addressing structural inequalities (Arabadjieva, 2016). Beyond Europe, the Inter-American Human Rights System has been leveraged by feminist NGOs in Brazil to address women's rights violations when national systems failed (Santos, 2018). Similarly, the Endorois community in Kenya used the African Commission on Human Rights to protect their land rights, overcoming political disadvantages at the national level (Gilbert, 2020).

Transnational litigation offers yet another avenue for disadvantaged groups to seek justice, particularly when domestic legal systems fail. Victims in developing countries frequently use foreign courts to hold multinationals accountable for human rights abuses (Meeran, 2021). For instance, South African asbestos miners turned to British courts to address corporate human rights violations (Boggio, 2006). In South Africa, environmental groups such as Earthlife Africa used strategic litigation to halt polluting projects and challenge government inaction, exemplified by the Life After Coal campaign, which also raised public awareness (Chamberlain and Fourie, 2024). These examples highlight how marginalized groups strategically navigate domestic, regional, and transnational courts to address systemic inequalities, leveraging judicial mechanisms to advance their rights.

2.2. Resource theory

Resource-based theory holds that success in legal mobilization depends significantly on available resources, both financial and legal knowledge and organization. This theory complements the theory of political disadvantage by focusing on the ability of actors, both individual and collective, to mobilize the law according to the resources available to them. Marc Galanter, in his influential article, argues that well-funded and organized litigators, referred to as "repeat players," are more likely to succeed in court than those without resources, due to their ability to invest time and money over the long term (Galanter, 1974). Vanhala reviews this approach, showing how the availability of resources remains a key factor in contemporary legal mobilization (Vanhala, 2017).

Notable examples -of use of resources in legal mobilisation at the national level- include cases like *Lustig-Prean and Beckett v. United Kingdom*, where gay rights organizations utilized fundraising and collaboration with NGOs to challenge discriminatory bans in the British armed forces (Mank and Smith, 2013). Similarly, in *S.L. v. Austria*, Austrian gay rights groups such as Rechtskomitee LAMBDA mobilized significant resources to abolish discriminatory laws, demonstrating how financial and legal support sustain successful campaigns (Mank and Smith, 2013).

In Latin America, climate litigation underscores the critical role of resources in advancing systemic change. In *Future Generations v. Colombia*, the Colombian State was ordered to repair climate damage, a ruling only achievable due to the plaintiff organizations' access to financial and legal resources (Barber, 2012). Conversely, a TWAIL (Third World Approaches to International Law) perspective highlights how resource disparities often disadvantage actors in the Global South. For instance, Bose (2023) critiques how human rights due diligence laws in France often exclude Global South voices due to resource barriers, while Buhmann (2023) shows that well-supported actors, such as those leveraging OECD National Contact Points, are more likely to succeed in corporate accountability cases.

Access to resources also influences success in regional litigation. For example, the Endorois and Ogiek indigenous peoples in Kenya overcame national barriers and secured justice through the African Commission on Human Rights, supported by NGOs like the Minority Rights Group (Claridge, 2019). Such examples reveal how financial and organizational assistance determines outcomes in international forums.

Finally, resource inequality shapes litigation dynamics in areas like climate justice and universal jurisdiction. Fossil fuel companies leverage their political and economic power to evade accountability, necessitating well-resourced litigation efforts to counteract their influence (Dehm, 2023). Similarly, Davis (2005) highlights how financial support is essential for prosecuting crimes under international law in national courts, while Hausler (2012)(2012) notes that Indigenous groups, like the Tsilhqot'in in Canada, rely heavily on NGO and legal aid for their legal mobilization efforts.

2.3. Theories of Legal Opportunities

The theory of legal opportunity structures highlights the importance of procedural rules that shape access to courts, determining who can litigate and under what conditions. Early research in political science focused on standing rules, showing

how these procedural mechanisms influence legal mobilization. Studies by Arrington (2019), Pedriana (2004) and Polletta (2000) demonstrated how activists leverage legal norms and processes, underscoring the critical role of accessible legal procedures in enabling mobilization.

Hilson (2002) expanded the concept of legal opportunity structures, linking procedural rules to public policy outcomes. Andersen (2006) further examined how litigants exploit opportunities created by courts. This framework has been applied in diverse contexts, including mobilization within the European Community (Evans Case and Givens, 2010) and Costa Rica (Wilson and Rodríguez Cordero, 2006). Vanhala's (2012) work illustrates how social movements can achieve procedural victories, even when substantive outcomes are unfavorable, demonstrating the strategic value of navigating legal opportunities.

Legal opportunity structures have driven corporate accountability in landmark cases like *Vedanta* and *Royal Dutch Shell* in the UK. British courts set precedents allowing parent companies to be held liable for the actions of their subsidiaries abroad, creating new pathways for victims of human rights violations to seek redress (Meeran, 2021). Similarly, in *Okpabi v. Shell*, Nigerian plaintiffs used UK legal mechanisms to overcome barriers in their home country, highlighting the role of transnational legal opportunities in advancing human rights (Roorda and Leader, 2021).

International mechanisms also offer vital legal opportunities for litigants. Sundstrom (2014) analyzed how Russian NGOs, constrained by domestic systems, used the European Court of Human Rights (ECtHR) to challenge human rights violations. Likewise, Wilson (2010) and Smart (2024) emphasized the transformative potential of regional courts like the Inter-American and African Courts of Human Rights, which empower marginalized groups by providing alternatives to weak national frameworks.

Domestic legal reforms can also expand opportunities for mobilization. Tiede (2011) examined judicial reforms in Chile, such as reduced pretrial detention and improved legal representation, which made the system more transparent and accessible. Yıldırım (2023), highlighted Turkey's "individual application" framework, which enables workers to directly access the Constitutional Court, advancing collective labor rights in a restrictive political environment. In Latin America, strategic litigation around climate change and gender justice has opened new procedural avenues for underrepresented groups. Urzola Gutiérrez (2024) argued that access to national and international tribunals is essential for the success of such initiatives, as it facilitates systemic changes in environmental and social justice contexts.

2.4. Social construction and framing

Social constructivist, framing, and translation approaches explain how social actors transform their demands into legal terms, attributing meanings to legal discourses and framing political objectives as plausible legal claims. Marshall's (2003) analysis of how women frame experiences of sexual harassment as human rights violations demonstrates how strategic framing impacts courts' reception of lawsuits. The framing of legal issues by individuals or groups directly influences the success of legal mobilization.

A clear example is *Goodwin v. United Kingdom*, where a transgender woman framed her demand for legal recognition under the right to privacy and family life,

leading to a favorable ruling (Anagnostou, 2010). Similarly, in *D.H. et al. v. Czech Republic*, litigants framed school segregation of Roma children as a violation of the rights to non-discrimination and education, a strategy pivotal to their success (Arabadjieva, 2016). In Latin America, Auz (2024) shows how climate justice cases reframe environmental harms as human rights violations under international treaties, broadening their legal impact.

Framing environmental litigation as a human rights issue is evident in *Earthlife Africa Johannesburg v. Minister of Environmental Affairs*, where activists linked coal plant construction to threats against environmental and human rights. This framing mobilized public opinion and delayed construction, demonstrating how strategic narratives can amplify litigation efforts (Chamberlain and Fourie, 2024). Similarly, Buhmann (2023) highlights how OECD National Contact Points are used not only as legal mechanisms but also as platforms to frame human rights claims.

Intersectional approaches further enrich framing strategies. De Jong (2024) argues for framing climate litigation to address intersecting oppressions, such as racism and sexism, enabling courts to assess human rights violations more comprehensively. Santos (2018) shows how feminist NGOs frame violations of women's rights within international human rights frameworks, facilitating transnational and domestic litigation when national systems fail.

In Canada, Cohen (2017) highlights how legal organizations framed maternal immunity debates using international children's rights standards, giving the issue broader judicial resonance. Hernandez and Levesque (2023), examine how the "I am a Witness" campaign framed racial discrimination against First Nations children as a fundamental human rights violation, increasing public and international pressure on the Canadian government.

Framing also plays a role in indigenous and counter-hegemonic litigation. Viaene and González-Serrano (2024) describe how indigenous lawyers in Guatemala frame their cases within their cultural identity and knowledge systems, challenging colonial and neoliberal structures. In torture and transnational crimes litigation, Mathews (2003) shows how framing claims as universal human rights violations enhances their success in international tribunals.

Murcott and Tigre (2024) emphasize the adaptability of framing in the Global South, where climate change litigation is situated within local contexts of justice and human rights. Finally, Kahle and Hammerslev (2023) illustrate how Indian NGOs frame their demands as "active citizenship," compelling the government to meet legal obligations despite reduced state responsibilities under neoliberal policies.

In conclusion, social construction, framing, and translation approaches are essential for understanding how legal narratives are crafted and deployed in diverse contexts. From environmental justice to gender rights, these strategies expand opportunities for legal mobilization, influencing outcomes in domestic, regional, and international courts.

2.5. Theories of Legal Consciousness

Legal consciousness refers to how individuals and groups shape their understanding of the law and how, in turn, the law influences their behaviors, values, and

identities. While widely studied in socio-legal research, it has received less attention in political science. Silbey's (2005) review of legal awareness and Ewick and Silbey's (1998) foundational work have provided theoretical frameworks that inform subsequent empirical studies on how marginalized actors perceive and use the law.

A critical aspect of legal consciousness is how peripheral groups view the law as a tool for resolving disputes and driving social change. Barber (2012) highlights how human rights organizations reflect critically on the law, recognizing its potential to influence public policy while acknowledging its limitations. Anzueto (2021), illustrates this in Guatemala, where strategic litigation not only secured reparations for human rights violations but also educated communities about their rights, fostering broader mobilization.

In Latin America, Cavallaro and Brewer (2008) emphasize the role of the Inter-American Court of Human Rights (IACHR) in fostering legal awareness. Landmark rulings have helped victims and human rights defenders view the court as a viable avenue for justice, enhancing perceptions of their rights and legal agency. Similarly, Smart (2024) notes how the IACHR's expansion of responsibility for human rights violations to include private companies has reshaped victims' perceptions of accountability, enabling litigation against corporations.

Legal mobilization in other regions reinforces the role of legal consciousness. Handmaker (2023) shows how strategic litigation in the Netherlands enhanced Roma communities' understanding of their rights, empowering them to use legal systems for protection. McGregor et al. (2019) argue that national human rights institutions (NHRIs) increase legal awareness through alternative mechanisms like mediation. However, they caution that such mechanisms should complement, not replace, access to courts, particularly in complex cases.

At the state level, Oette (2012) examines how national legal cultures shape the implementation of anti-torture laws, with stronger legal systems more effectively integrating international standards. In Chile, Tiede (2011), observes that judicial reforms have increased transparency and trust, shifting public perceptions of justice and rights. Urzola Gutiérrez (2024) adds a feminist perspective, arguing that marginalized groups, especially women, construct legal awareness through distinct epistemologies, challenging dominant narratives in climate litigation.

In summary, legal consciousness provides a lens for understanding how individuals and groups interact with the legal system and how these interactions shape their perceptions of justice. From grassroots mobilization to international litigation, it highlights the reciprocal relationship between law and society, demonstrating the law's role as both a tool and a site of social change.

3. Litigation against AI companies

Having analyzed the theoretical approaches to legal mobilization, we now turn to examine their practical application in the context of AI and potential human rights abuses. This next section explores the types of litigation pursued, the jurisdictions where these cases have been brought, and the framing strategies employed by litigants. By delving into specific examples, we aim to illustrate how the concepts of political disadvantage, resource mobilization, legal opportunities, framing, and legal

consciousness manifest in addressing the challenges posed by AI in human rights contexts.

To carry out this analysis, we have extracted the cases from two key databases:

a) The Tech Litigation Database: This open-access resource, hosted at the *Centre for Fundamental Rights* at the *Hertie School* in Berlin, collects, explores, and exposes the growing number of global litigations against automated systems. Its purpose is to raise awareness and provide transparency about the impact of new technologies on society, supporting the work of researchers, legal actors and civil society organizations.

b) DAIL – Database of AI Litigation: This database, maintained by *George Washington University*, presents information on ongoing and finalized litigation involving AI, including machine learning. It covers cases from the beginning of the lawsuit, ranging from algorithms used in hiring decisions, credit and criminal sentences, to liability for accidents with autonomous vehicles. In addition, it includes some cases on statistical analysis and data protection that, although they do not directly involve AI, are of relevance for AI projects.

Our analysis will be based on these 50 cases and will allow us to gain a global view of how courts and other legal actors are responding to the use of AI technologies in various application areas. We will also explore how these cases can be theorized under the framework of strategic legal mobilization, as addressed in the previous sections of the chapter.

3.1. What kind of trials are taking place?

In recent years, we have seen a growing variety of lawsuits addressing the legal implications of using AI and other automated technologies. These cases cover a wide range of application areas, from intellectual property rights to the right to privacy and non-discrimination. Through the analysis of the 50 cases, we can identify several types of trials that mark the beginning of a new era in strategic litigation around these technologies.A significant first group of lawsuits revolves around data protection and privacy, a central issue in most GAI-related litigation. A key example is the case of the German Bank, where Berlin's *Commissioner for Data Protection and Freedom of Information* fined a bank €300,000 for failing to properly inform its customers about the criteria used in granting or refusing credit cards, in violation of the European Union's General Data Protection Regulation (GDPR).[3] This case demonstrates how the use of AI in financial decisions requires a level of transparency that many companies are not yet meeting.

In a related case, the Italian Data Protection Authority imposed a €20 million fine on a U.S. company for the illegal use of AI-based biometric monitoring technologies. The company had collected biometric data from more than 10 billion public images without proper consent, thus violating the principles of purpose limitation and

[3] Berlin Commissioner for Data Protection and Freedom of Information. (2024). *Decision on automated credit card issuing*.

transparency set forth in the GDPR.[4] These cases underscore the importance of consent and transparency in the use of AI, especially when it comes to sensitive data.

Another type of judgment that stands out concerns accountability and transparency in the use of automated systems for public administration. For example, in the SyRI (System Risk Indication) case in the Netherlands, the court found that the use of AI to detect fraud in the social welfare system violated the right to privacy under the *European Convention on Human Rights*.[5] The court ruled that the system was not transparent enough and that the automated decision-making process did not comply with the principles of fairness and proportionality.

Judgments are also underway around algorithmic discrimination and bias in automated decisions. In the case of R (Bridges) v. Chief Constable of South Wales Police, a UK court found that the use of facial recognition technologies by police violated the right to privacy and lacked adequate review to prevent racial or gender bias.[6] This ruling reflects growing scrutiny on how AI systems can perpetuate or even amplify structural discrimination.

In the area of intellectual property and patents, the Tokyo District case highlighted how courts are grappling with the question of whether AI can be recognized as inventors. In this case, the patent application filed by DABUS, an AI system, was rejected on the basis that patents can only cover human inventions.[7] The decision underscores the need for a more robust legal framework that addresses the complexities of invention in the age of AI.

As for corporate liability, the case of Prygodicz v. Commonwealth of Australia, known as the "Robodebt" case, refers to an automated system used by the Australian government to recover debts from the social security system.[8] The federal court found that the system was unlawful because it shifted the burden of proof onto beneficiaries, requiring them to prove they didn't owe money, rather than the government proving otherwise. This case exemplifies how AI, when implemented without sufficient oversight, can have devastating consequences for citizens' rights.

On the other hand, we also find judgments that address the interaction between AI and labor rights. In the case of Rb. Amsterdam, drivers of a taxi app challenged the deactivation of their accounts based on fully automated decisions, arguing that the lack of human intervention violated their right to work and due process.[9] The court ordered the company to reinstate the drivers and highlighted the need for greater transparency and human oversight in AI-powered employment decision-making.

In addition, lawsuits related to the use of facial recognition and other biometric technologies are also prominent. In a case in Kenya, the court declared that the digital identification system based on facial recognition, known as Huduma Namba,

[4] Italian Data Protection Authority, *Decision no. 50 of 10.02.2022*, regarding AI-based biometric monitoring technologies.

[5] District Court of The Hague, *SyRI (System Risk Indication)*, ECLI:NL:RBDHA:2020:187 (5 February 2020).

[6] R (Bridges) v Chief Constable of South Wales Police [2020] EWCA Civ 1058.

[7] Tokyo District Court, *DABUS Patent Application*, R5 (9gyo-u) no. 5001 (16 May 2024).

[8] Prygodicz v Commonwealth of Australia [2021] FCA 941.

[9] Rb. Amsterdam - C/13/684665/KG ZA 20-481 (2020).

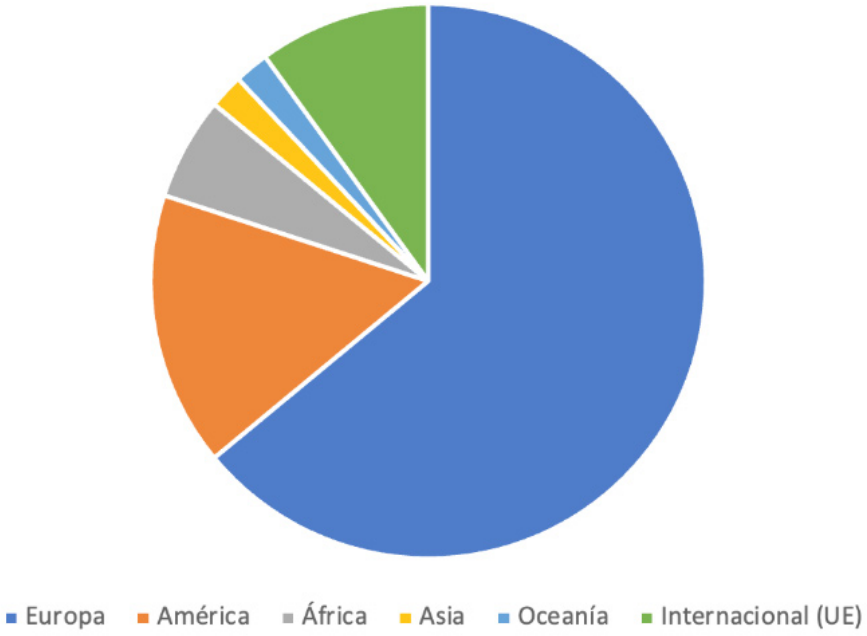
violated data protection laws, requiring the government to pass more stringent laws before continuing its use.[10]

These examples demonstrate that trials around AI not only cover a wide range of application areas but are also testing the limits of traditional legal frameworks. At the same time, these cases reflect how fundamental rights, such as privacy, non-discrimination, and transparency, are being reinterpreted and reinforced in a world where AI plays an increasingly dominant role.

3.2. In which jurisdictions?

The analysis of the 50 cases extracted from The *Tech Litigation Database* and *DAIL* databases shows that the jurisdictions where most of the litigation against AI companies is taking place are diverse, covering countries in Europe as well as America, Asia and Africa (see Figure 1). However, some jurisdictions play a more prominent role due to the magnitude and complexity of the cases, as well as the impact of their decisions on the development of the legal framework that regulates these technologies.

Figure 1: Jurisdictions of 50 analysed cases



Source: own creation

Europe stands out as one of the key regions in litigation against GAI, in part thanks to the existence of the *General Data Protection Regulation* (GDPR), which establishes a rigorous framework for data protection and privacy. A paradigmatic example is the case of the SyRI (System Risk Indication) System in the Netherlands, where the District Court of The Hague ruled that the use of algorithms to detect fraud in the social welfare system violated the right to privacy under Article 8 of the *Euro-*

[10] Nubian Rights Forum & 2 others v. Attorney General & 6 others [2020] eKLR.

pean Convention on Human Rights (ECHR).[11] This case underlines the importance of the balance between automated technology and fundamental rights in European jurisdictions.

Germany is also emerging as a crucial jurisdiction in protecting citizens' rights from AI systems. In a particular case the German Federal Court examined the use of an automated credit scoring system, ruling that decisions affecting individuals must include human supervision to ensure transparency and access to information.[12] This case is representative of the rigorous approach of the German courts in ensuring that consumers' rights are not compromised by automated decisions.

Italy is another notable European jurisdiction, in a particular case the Italian Supreme Court declared invalid the consent of users to an automated system that affected their rights if they were not provided with adequate information on the operation of the algorithm.[13] This case underscores how Italian courts are promoting transparency and the protection of individual rights in the use of AI technologies.

In addition to Europe, the United States is presented as another priority jurisdiction. In the case of State of New Jersey v. Francisco Arteaga, the New Jersey Court of Appeals ruled that the prosecutor must provide detailed information about facial recognition technology used to identify a suspect, as denying this access would infringe on the right to due process.[14] This case demonstrates the focus of U.S. courts on ensuring that new technologies do not compromise the procedural rights of individuals.

Another significant example in the United States is the case of State v. Loomis, in which the Wisconsin Supreme Court upheld the use of risk assessment algorithms in sentencing decisions, but emphasized the need to include written warnings about the algorithm's limitations to ensure fairness.[15] This ruling reflects how U.S. courts are grappling with the use of AI in the criminal justice system and the need for a balance between technological efficiency and constitutional rights.

In Latin America, Brazil has been an important player, especially in the context of privacy and data protection rights. The IDEC v. Via Quatro case in Brazil addressed the use of facial recognition by a public transport company without proper consent from users, resulting in significant compensation for moral damages. This case illustrates how Brazilian courts are beginning to address the impact of AI on citizens' privacy. Africa also stands out with important jurisdictions such as Kenya, where in the case of Nubian Rights Forum already analysed.[16] And in Asia, Japan has begun to see significant judgments in this area.[17]

Finally, cases before the Court of Justice of the European Union (CJEU), have shown that decisions on the retention of personal data and the use of AI in Europe can have a large-scale impact, influencing the regulation and use of technologies across

[11] District Court of The Hague, *SyRI (System Risk Indication)*, ECLI:NL:RBDHA:2020:187 (5 February 2020).

[12] BGH VI ZR 156/13 (Schufa) (28 January 2014).

[13] Corte Suprema di Cassazione, Civil Ord., Sez. 1, No. 14381 (2023).

[14] State of New Jersey v Francisco Arteaga, New Jersey Court of Appeals (2023).

[15] State v Loomis, 881 N.W.2d 749 (Wis. 2016).

[16] Nubian Rights Forum & 2 others v. Attorney General & 6 others [2020] eKLR.

[17] See for example Tokyo District Court. (2024). *DABUS Patent Application*, R5 (9gyo-u) no. 5001.

the continent.[18] These cases highlight how the European legal framework is adapting to new technological realities.

In summary, the priority jurisdictions where litigation against AIcompanies is taking place cover both Europe and the Americas, Asia and Africa, with Europe at the forefront due to the strength of its data protection and human rights legal framework. At the same time, the United States, Brazil, Kenya, and Japan are emerging as key players in the evolution of regulation and litigation around AI. These cases are setting important precedents and helping to define how these technologies will be used and regulated in the future.

3.3. What are the predominant causes and human rights in litigation?

In analyzing the 50 cases drawn from *The Tech Litigation Database* and *DAIL,* we observed several preponderant causes of action ranging from the protection of personal data to the misuse of AI technologies in areas such as facial recognition, digital privacy, and discrimination. The causes of action are concentrated in five broad areas: privacy violations, data protection, discrimination, violations of the rights of access to information, and failures in transparency and human oversight (see figures 2 and 3).

3.3.1. Preponderant Causes of Action

Data protection and privacy: One of the main causes of action is related to the violation of the right to privacy and data protection, mainly regulated by the *General Data Protection Regulation* (GDPR) in Europe. Cases such as SyRI (System Risk Indication) in the Netherlands show how the use of algorithms to detect fraud in social welfare systems was declared illegal for violating citizens' right to privacy under Article 8 of the *European Convention on Human Rights* (ECHR). In another example, in Canada, an AI company was found guilty of collecting and using biometric data without consent, leading to an order to stop the collection and delete the stored data.[19]

Algorithmic discrimination: Discrimination, both racial and gender, is another prominent cause of action in litigation against AGI systems. In Italy, a food delivery platform was fined for using algorithms that favored certain workers over others, based on a scoring system that ended up excluding many delivery drivers, exacerbating discrimination and affecting their right to work.[20] Similarly, in the United Kingdom, a Court found that the mass retention of immigrants' mobile data without proper procedures constituted undue discrimination.[21]

Lack of transparency and human oversight: Several causes of action are related to the lack of transparency in the use of AI technologies and the lack of adequate human oversight in decision-making processes. An example from Austria, shows how a court found that a credit rating company did not provide enough information about the

[18] See for example CJEU, Cases C-793/19 and C-794/19, *La Quadrature du Net and Others v France; Ordre des barreaux francophones et germanophone and Others v Belgium* [2020] ECLI:EU:C:2020:791.

[19] Office of the Privacy Commissioner of Canada, *PIPEDA Findings #2021-001.*

[20] Italian Data Protection Authority, *Decision No. 9675440*, regarding algorithmic discrimination in the food delivery sector (2021).

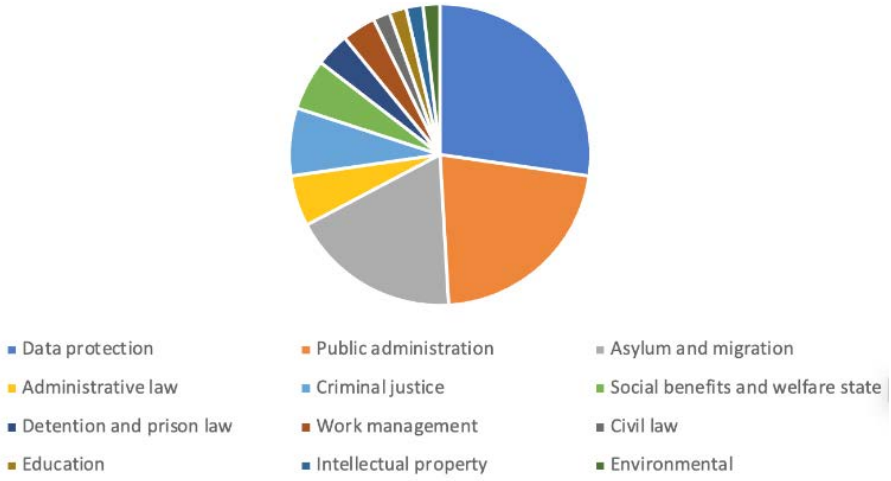
[21] HM, MA, and KH v Secretary of State for the Home Department [2023] EWHC (Admin).

logic behind automated decisions, which violated Article 15 of the GDPR.[22] Another case is the already cited case of State v. Loomis in the United States, where the court confirmed that the use of a risk assessment algorithm in the conviction of a defendant required clear warnings about the limitations of the system to ensure a fair trial.[23]

Access to information and transparency of algorithms: Another focus of litigation is access to information, especially when it comes to algorithms used in automated decisions. The case of the Netherlands, where drivers of a taxi app sued for a lack of transparency about the reasons for the deactivation of their accounts, is a representative example.[24] The court determined that the deactivation constituted an automated decision and ordered the reinstatement of the drivers, highlighting the importance of access to the logic behind algorithmic decisions.

Social benefits: Many of the causes of action also arise from violations of social rights, as in the case of Kammarrätten i Göteborg in Sweden, where a court ruled that the denial of access to an algorithm that managed social benefits constituted a violation of access to information rights under the Swedish constitution.[25]

Table 2: Preponderant causes of action



Source: Own creation

3.3.2. Preponderant Human Rights in Litigation

In terms of overriding human rights, litigation against AI companies tends to focus on several fundamental rights, mainly those related to privacy, non-discrimination, access to information and the right to a fair trial.

Right to privacy: This is one of the most affected rights in litigation related to GAI, especially in those cases involving the use of facial recognition technologies and data extraction without consent. In Glukhin v Russia, the European Court of Human Rights (ECtHR) found that the Russian government's use of facial recognition technology to identify a protester violated his right to privacy under Article 8 of the

[22] Vienna Regional Administrative Court, *VGW-101/042/791/2020-44* (2020).

[23] State v Loomis, 881 N.W.2d 749 (Wis. 2016).

[24] Rb. Amsterdam - C/13/684665/KG ZA 20-481 (2020).

[25] Kammarrätten i Göteborg, *Mål nr 6973-19*, 21 February 2020.

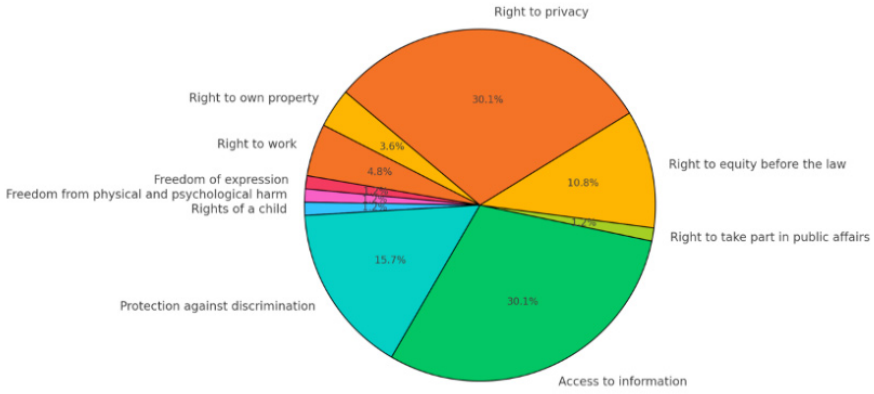
ECHR.[26] Similarly, in the GPS tracking pilot case in the UK, the Information Commissioner's Office ruled that the continuous collection of data without proper consent violated the right to privacy of migrants monitored using GPS devices.[27]

Right to non-discrimination: Algorithmic discrimination is another crucial aspect, and it is found in cases where the use of AI has exacerbated pre-existing biases. In Spain, for example, a supermarket was sanctioned for using a facial recognition system that discriminated against individuals based on criminal records, violating the principles of necessity and proportionality of the GDPR.[28] Also, the already analysed case of Schufa in Germany, where the court addressed how credit scoring systems can discriminate against certain individuals, particularly those with low incomes.[29]

Right to information and transparency: Access to information is a key right in many of the AI cases, especially those where algorithmic decisions significantly impact individuals. In La Quadrature du Net in France, the court annulled an agreement between a regional council and a technology company due to the lack of clear consent and proportionality in the use of facial recognition systems in schools, highlighting the right of students and their families to be properly informed about the use of their personal data.[30]

Right to a fair trial and due process: Fairness in the use of AGI's technologies in the judicial system has also been a prominent concern. In State v. Loomis in the United States, the Wisconsin Supreme Court emphasized the need to provide clear information on the limitations of the risk assessment algorithms used in sentencing, to ensure that the right to due process is not violated.[31] Similarly, in Sigurour Einarsson and others v Iceland, the lack of equitable access to digital evidence was questioned, which affected the right of the accused to a fair trial.[32]

Figure 3: Human rights impacted



Source: own creation

[26] Glukhin v Russia (Application no. 11519/20), ECtHR (2024).

[27] Information Commissioner's Office, *GPS Tracking Pilot Case*, United Kingdom (2024).

[28] Spanish Data Protection Agency, *Decision PS/00120/2021* (2021).

[29] BGH VI ZR 156/13 (Schufa) (28 January 2014).

[30] Administrative Court of Marseille, *La Quadrature du Net v Provence-Alpes-Côte d'Azur Regional Council* (2020).

[31] State v Loomis, 881 N.W.2d 749 (Wis. 2016).

[32] European Court of Human Rights. (2021). *Sigurður Einarsson and Others v. Iceland* (Application no. 39757/15).

In summary, the predominant causes of action in litigation against AIcompanies mainly include privacy violations, discrimination, lack of transparency, access to information and failures in human supervision. Meanwhile, the most commonly affected human rights are the right to privacy, non-discrimination, access to information, the right to a fair trial, and the right to work. These litigations are helping to set important precedents that will guide the development of regulatory frameworks for AI in the future.

4. As a mode of conclusion: bridging the gap between theory and litigation against AI companies

By looking at the cases analyzed in AI and neurotechnologies, a clear connection can be established with the theories of strategic legal mobilization discussed above. Theories that explore how social and collective actors turn to the courts to advance in the defense of their rights find direct resonance in these emerging litigations against AI companies. Through the application of these theories, it is possible to develop a conceptual framework that allows us to understand not only the motives behind these cases, but also the strategies used by litigants to confront inequalities caused or exacerbated by technology.

One of the theories that becomes relevant when reviewing cases is the theory of political disadvantage. Many of the actors who have turned to the courts to litigate AI companies represent marginalized or vulnerable groups, who lack political power or access to traditional decision-making mechanisms. A clear example is a case in Germany, in which an asylum seeker faced data extraction and analysis without due privacy guarantees.[33] In this context, legal mobilization before the courts becomes the only avenue available to protect the rights of those who face structural disadvantages vis-à-vis state or corporate actors.

In line with this theory, many cases show how litigants turn to international or supranational courts when national legal structures are not sufficient to provide effective protection. The case *of Glukhin v. Russia*, where the European Court of Human Rights (ECHR) ruled against the use of facial recognition on protesters, demonstrates how politically marginalized actors turn to external judicial systems to assert their rights.[34] This type of mobilization is particularly relevant in contexts where human rights, such as privacy and freedom of expression, are systematically violated at the national level.

The resource-based theory is also essential to understand these disputes. As previously discussed, access to financial, legal, and organizational resources can be determinant in the success of legal mobilization. Several cases reflect how litigants who manage to mobilize sufficient resources can sustain protracted litigation against large companies. In the *PIPEDA* in Canada, a company was sanctioned for the misuse of biometric data, but it was only possible due to access to the necessary resources by the regulatory authority to sustain an in-depth investigation.[35] Likewise, the IT *DPA Decision* case in Italy, which involved a food delivery platform, highlights how hu-

[33] Administrative Court of Berlin, *VG 9 K 135/20 A* (2020).

[34] Glukhin v Russia (Application no. 11519/20), ECtHR (2024).

[35] Office of the Privacy Commissioner of Canada, *PIPEDA Findings #2021-001*.

man and labour rights organisations have played an essential role in providing the necessary resources to take these litigations to court.[36]

These cases also show how actors' resort to international legal mobilization when they lack access to sufficient resources at the national level. This aligns with criticisms of TWAIL approaches, which point out how litigants in the Global South often rely on international actors to mobilize effectively in court. Cases such as *Kammarrätten i Göteborg* in Sweden, where a journalist requested access to the source code of an algorithm used in the administration of social benefits, are examples of how inequalities in access to technological and legal resources can determine the outcome of litigation.[37]

Legal opportunity theory is also particularly useful for understanding how strategic litigants have taken advantage of procedural gaps and opportunities to advance their claims. In the State *v. Loomis* case in the United States, the courts examined how an automated risk assessment system was used in the conviction of a defendant. Although the court allowed the use of the algorithm, the case highlighted the need to review the legal opportunities available to protect due process rights. This decision underscores how procedural opportunities, in this case the right to understand the logic behind automated decisions, can become the key point of strategic litigation.

Access to the courts of foreign jurisdictions also reflects how legal opportunities expand beyond national frameworks. Cases such as *SyRI* in the Netherlands, where an algorithm was declared illegal for violating the right to privacy, underscore the importance of procedural structures that allow litigants to challenge the practices of governments or companies in other jurisdictions. This ability to litigate in different jurisdictional contexts is a clear reflection of how legal opportunity structures have evolved, allowing actors to access more favorable judicial mechanisms.

In addition, theories of social construction and framing play a crucial role in how demands are formulated and presented in these cases. In litigation such as *La Quadrature du Net* in France, framing facial recognition as a threat to fundamental rights allowed plaintiffs to construct a powerful narrative in court. The way in which social actors frame their claims, whether in terms of privacy violation or algorithmic discrimination, can directly influence the reception these claims have in court. In this sense, litigation against AIcompanies is a clear example of how legal actors construct narratives that seek to amplify the relevance of human rights in the context of technology.

Litigants in these cases not only seek favorable outcomes in legal terms, but also attempt to influence public opinion and policy through their strategic frameworks. Cases such as *Bridges v. CC South Wales* in the United Kingdom, which challenged the use of facial recognition technologies by police, illustrate how litigants are attempting to transform the debate about privacy and fundamental rights, not only in the courts, but also in the public arena. These types of framing strategies, in combination with legal mobilization, allow the demands around emerging technologies to gain greater visibility.

[36] Italian Data Protection Authority, *Decision No. 9675440*, regarding algorithmic discrimination in the food delivery sector (2021).

[37] Kammarrätten i Göteborg, *Mål nr 6973-19*, 21 February 2020.

On the other hand, legal awareness theory is useful for analyzing how litigants and affected groups develop a critical understanding of their rights in relation to technology. In the case *of Sigurour Einarsson and others v. Iceland*, the lack of equitable access to data used by the prosecution led to a discussion about procedural fairness in the use of advanced technologies. As litigants begin to understand how GAI's technologies can influence their rights, they develop a legal awareness that allows them to challenge surveillance and automation practices in court.

The development of critical legal awareness is also present in cases such as *Nubian Rights Forum v. Attorney General* in Kenya, where litigants questioned the implementation of a digital ID system without adequate safeguards for the protection of personal data. These litigations reflect how actors affected by emerging technologies not only seek to remedy a specific situation, but also develop a broader awareness of how the right can be used to protect their rights in the face of new forms of digital surveillance.

Finally, litigation around AI and neurotechnology is also influencing the development of more robust regulatory frameworks for the protection of human rights. The cases analysed, such as *Automated credit card issuing* in Germany, underline how strategic legal mobilisation not only seeks to obtain individual compensations, but also drives structural reforms. By taking these cases to court, legal actors are generating precedents that can serve as the basis for future regulations that address the challenges of technology in protecting human rights.

In summary, the cases litigated against AI companies not only illustrate how theories of strategic legal mobilization can be applied in practice, but also demonstrate how social actors are using the courts to address the inequalities and rights violations that arise at the intersection of technology and human rights. These litigations are laying the groundwork for new forms of legal protection in the face of technological advances, and theories of strategic legal mobilization provide a solid framework for understanding how and why these mobilizations are occurring.

References

Anagnostou D (2010) Does European human rights law matter? Implementation and domestic impact of Strasbourg Court judgments on minority-related policies. *The International Journal of Human Rights* 14(5). Routledge: 721–743.

Andersen EA (2006) *Out of the Closets and into the Courts*: Legal Opportunity Structure and Gay Rights Litigation. Ann Arbor: University of Michigan Press.

Anzueto M-A (2021) In the shadow of Canadian imperialism? Strategic human rights litigation in Guatemala (2009–2019). *The International Journal of Human Rights* 25(10). Routledge: 1659–1681.

Arabadjieva K (2016) Challenging the school segregation of Roma children in Central and Eastern Europe. *The International Journal of Human Rights* 20(1). Routledge: 33–54.

Arrington CL (2019) Hiding in Plain Sight: Pseudonymity and Participation in Legal Mobilization. *Comparative Political Studies* 52(2): 310–341.

Auz J (2024) The Political Ecology of Climate Remedies in Latin America and the Caribbean: Comparing Compliance between National and Inter-American Litigation. *Journal of Human Rights Practice* 16(1): 182–207.

Barber CC (2012) Tackling the evaluation challenge in human rights: assessing the impact of strategic litigation organisations. *The International Journal of Human Rights* 16(3). Routledge: 411–435.

Boggio A (2006) The Global Enforcement of Human Rights: The Unintended Consequences of Transnational Litigation. *The International Journal of Human Rights* 10(4). Routledge: 325–340.

Bose D (2023) Decentring Narratives around Business and Human Rights Instruments: An Example of the French *Devoir de Vigilance* Law. *Business and Human Rights Journal* 8(1): 18–42.

Buhmann K (2023) Confronting Challenges to Substantive Remedy for Victims: Opportunities for OECD National Contact Points under a Due Diligence Regime Involving Civil Liability. *Business and Human Rights Journal*: 1–24.

Cavallaro JL and Brewer SE (2008) Reevaluating Regional Human Rights Litigation in the Twenty-First Century: The Case of the Inter-American Court. *The American Journal of International Law* 102(4). American Society of International Law: 768–827.

Chamberlain L and Fourie M (2024) Using Climate Litigation to Strengthen Advocacy Strategies: The Life After Coal Campaign in South Africa. *Journal of Human Rights Practice* 16(1): 248–257.

Claridge L (2019) The approach to UNDRIP within the African Regional Human Rights System. *The International Journal of Human Rights* 23(1–2). Routledge: 267–280.

Cohen M (2017) International human rights norms and maternal tort immunity in Canada: connecting the dots. *The International Journal of Human Rights* 21(6). Routledge: 655–671.

Cortner RC (1968) Strategies and Tactics of Litigants in Constitutional Cases. *Journal of Public Law* 17(2): 287–307.

Davis M (2005) Universal Jurisdiction: National Courts and the Prosecution of Serious Crimes under International Law (review). *Human Rights Quarterly* 27(2). Johns Hopkins University Press: 729–735.

de Jong IJM (2024) Beyond the turn to human rights: a call for an intersectional climate justice approach. *The International Journal of Human Rights* 28(5). Routledge: 738–758.

Dehm J (2023) Beyond Climate Due Diligence: Fossil Fuels, 'Red Lines' and Reparations. *Business and Human Rights Journal* 8(2): 151–179.

Epstein L (1985) *Conservatives in Court*. University of Tennessee Press.

Evans Case R and Givens TE (2010) Re-engineering Legal Opportunity Structures in the European Union? The Starting Line Group and the Politics of the Racial Equality Directive. *JCMS: Journal of Common Market Studies* 48(2): 221–241.

Ewick P and Silbey SS (1998) *The Common Place of Law: Stories from Everyday Life*. Chicago Series in Law and Society. Chicago, IL: University of Chicago Press. Available at: https://press.uchicago.edu/ucp/books/book/chicago/C/bo3615912.html (accessed 21 October 2024).

Galanter M (1974) Why the Haves Come out Ahead: Speculations on the Limits of Legal Change Essay. *Law & Society Review* 9(1): 95–160.

Gilbert J (2020) Indigenous Peoples and Litigation: Strategies for Legal Empowerment. *Journal of Human Rights Practice* 12(2): 301–320.

Handmaker J (2023) Introduction to the legal mobilization special focus. *Journal of Human Rights Practice* 15(1): 1–5.

Hausler K (2012) Indigenous perspectives in the courtroom. *The International Journal of Human Rights* 16(1). Routledge: 51–72.

Hernandez J and Levesque A (2023) Movement Lawyering and the *Caring Society* Litigation. *Journal of Human Rights Practice* 15(2): 395–413.

Hilson C (2002) New social movements: the role of legal opportunity. *Journal of European Public Policy* 9(2): 238–255.

Kahle A and Hammerslev O (2023) Activating Citizenship through NGO-Led Litigation: Shaping the Neoliberal State to Eradicate Manual Scavenging in India. *Journal of Human Rights Practice* 15(2): 414–431.

Ligthart S, Bublitz C and Alegre S (2023) Neurotechnology: We need new laws, not new rights. *Nature* 620: 950.

Mank BC and Smith S (2013) The Environmental Rights Revolution: A Global Study of Constitutions, Human Rights, and the Environment by David R. Boyd (review). *Human Rights Quarterly* 35(4). Johns Hopkins University Press: 1021–1042.

Marshall A-M (2003) Injustice Frames, Legality, and the Everyday Construction of Sexual Harassment. *Law & Social Inquiry* 28(3): 659–689.

Mathews S (2003) Torture as Tort: Comparative Perspectives on the Development of Transnational Human Rights Litigation (review). *Human Rights Quarterly* 25(1). Johns Hopkins University Press: 267–274.

McGregor L (2012) The role of supranational human rights litigation in strengthening remedies for torture nationally. *The International Journal of Human Rights* 16(5): 737–754.

McGregor L, Murray R and Shipman S (2019) Should National Human Rights Institutions Institutionalize Dispute Resolution? *Human Rights Quarterly* 41(2). Johns Hopkins University Press: 309–339.

Meeran R (2021) Multinational Human Rights Litigation in the UK: A Retrospective. *Business and Human Rights Journal* 6(2): 255–269.

Murcott MJ and Tigre MA (2024) Developments, Opportunities, and Complexities in Global South Climate Litigation: Introduction to the Special Collection. *Journal of Human Rights Practice* 16(1): 1–24.

Oette L (2012) Implementing the prohibition of torture: the contribution and limits of national legislation and jurisprudence. *The International Journal of Human Rights* 16(5). Routledge: 717–736.

Pedriana N (2004) Help Wanted NOW: Legal Resources, the Women's Movement, and the Battle Over Sex-Segregated Job Advertisements. *Social Problems* 51(2): 182–201.

Polletta F (2000) The Structural Context of Novel Rights Claims: Southern Civil Rights Organizing, 1961-1966 Constitutional Interpretation and Institutional Analysis. *Law & Society Review* 34(2): 367–406.

Roorda L and Leader D (2021) Okpabi v Shell and Four Nigerian Farmers v Shell: Parent Company Liability Back in Court. *Business and Human Rights Journal* 6(2): 368–376.

Santos CM (2018) Mobilizing Women's Human Rights: What/Whose Knowledge Counts for Transnational Legal Mobilization? *Journal of Human Rights Practice* 10(2): 191–211.

Silbey SS (2005) AFTER LEGAL CONSCIOUSNESS. *Annual Review of Law and Social Science* 1(1): 323–368.

Smart S (2024) Expanding and Contracting the UN Guiding Principles: an Analysis of Recent Inter-American Human Rights Court Decisions. *Journal of Human Rights Practice* 16(1): 342–354.

Smart S and Oyarzun E (2024) El caso Girardi c. EMOTIV Inc. ante la Corte Suprema de Chile: ¿un paso adelante en la protección de los neuroderechos humanos? In: *Desafios Da Interface Neurodireito e Inteligência Artificial*. Brazil: Livraria do Avdogado.

Sundstrom LM (2014) Russian NGOs and the European Court of Human Rights: A Spectrum of Approaches to Litigation. *Human Rights Quarterly* 36(4). Johns Hopkins University Press: 844–868.

Tiede LB (2011) Taking rights seriously in Chile. *The International Journal of Human Rights* 15(8). Routledge: 1275–1290.

Urzola Gutiérrez N (2024) Gender in Climate Litigation in Latin America: Epistemic Justice Through a Feminist Lens. *Journal of Human Rights Practice* 16(1): 208–226.

Vanhala L (2012) Legal Opportunity Structures and the Paradox of Legal Mobilization by the Environmental Movement in the UK. *Law & Society Review* 46(3): 523–556.

Vanhala, L (2017) Is Legal Mobilization for the Birds? Legal Opportunity Structures and Environmental Nongovernmental Organizations in the United Kingdom, France, Finland, and Italy. *Comparative Political Studies* 51(3): 380–412.

Viaene L and González-Serrano MX (2024) The right to be, to feel and to exist: Indigenous lawyers and strategic litigation over Indigenous territories in Guatemala. *The International Journal of Human Rights* 28(4). Routledge: 555–577.

Vose CE (1959) Caucasians Only: The Supreme Court, the NAACP, and the Restrictive Covenant Cases. Univ of California Press.

Wilson BM and Rodríguez Cordero JC (2006) Legal Opportunity Structures and Social Movements: The Effects of Institutional Change on Costa Rican Politics. *Comparative Political Studies* 39(3): 325–351.

Wilson RJ (2010) Regional Protection of Human Rights (review). *Human Rights Quarterly* 32(1). Johns Hopkins University Press: 227–234.

Yıldırım E (2023) Collective labour rights and the Turkish constitutional court. *The International Journal of Human Rights* 27(4). Routledge: 734–753.

Yuste R (2023) Advocating for neurodata privacy and neurotechnology regulation. *Nature Protocols* 18(10). Nature Publishing Group: 2869–2875.

— 11 —

The inter-American region is regulating the use of neurotechnologies. What are the strategies of democratic systems?[1] [2]

NATALIA LEONOR MONTI
Fundación Kamanau
nmonti@gmail.com

Introduction

In recent times, there have been warnings from different sectors, especially from the laboratories that research and put into practice neurotechnological techniques, that these advances are incredibly innovative and can significantly improve people's quality of life. However, they can also jeopardize various manifestations of legal guarantees, perhaps until now imperceptible from the classical and traditional viewpoint of law.

This is why it is essential to focus on a paradigm shift. There is a need to protect democratic systems in the face of new crossroads imposed by the scientific advances of neurotechnologies and their convergence with AI and immersive technologies; we must balance innovation with the protection of citizens' rights, because it affects the functioning of democracy (Monti, Natalia; Sánchez, Moisés & Colombara, Ciro, 2023).

In this sense, the news announced by the company Neuralink, a few months ago, about the brain implant in a quadriplegic patient that allows him to play online chess by moving the mouse only with his mind, has reached great repercussion (Neuralink, 2024). This could have a considerably larger number of practical applications, which are also being explored by other communications and technology giants, such

[1] I thank Moisés Sánchez for his critical view and conceptual contributions to this document.

[2] This document was translated in collaboration with Deepl's Artificial Intelligence.

as Meta, Microsoft and Google. Thanks to the competition among these companies to get first and best to market with a device that allows us to manipulate a phone, video game or other device using only our mind, it is already likely that we will soon be able to interact with our environment without using our hands or our voice. We will have the technology built into our brains to do so. At least, this is what the companies that are developing these devices claim. For example, in recent times we have observed how Neurogaming has been growing: playing with the mind. What seemed only science fiction is now a reality and there are those who play video games only with their brains. That is, without the need to manipulate a joystick (Playing Video Games With Mind Control, 2023).

A new reality is complex if the previous regulatory precautions are not taken, due to the influence that these developments can have on democratic systems, through manipulation in decision-making, invasion of privacy and the affectation of electoral systems. It is also necessary to protect identity and agency, the latter understood in its sociological sense: the ability to choose one's actions with free will. On the other hand, there is concern about providing greater protection to vulnerable groups, due to various factors, including the possibility of "artificial augmentation" of brain capacities (which could produce serious inequalities).

In this context, a movement was generated that began to promote the need to establish regulations to protect people from certain threats that were observed with the advance of neuroscientific developments.

In recent years, various initiatives have been developed to minimize the impact of neurotechnology applications. Mentioning that recently in our Inter-American region, significant progress was made with the first declaration of its kind worldwide, since we have the development of international standards with the work prepared by the Inter-American Juridical Committee (CJI) of the OAS (Comité Jurídico Interamericano, 2023).

In addition, the international environment is also advancing in important processes for the generation of global standards. This is the case of UNESCO, which has initiated work to develop a global Recommendation on the ethics of neurotechnology (UNESCO, Towards an International Instrument, 2025) through various regional consultations with local experts (UNESCO, 2024).

Furthermore, the UN Human Rights Council, in its resolution 51/3, requested the Advisory Committee to prepare a study on the effects, opportunities and challenges of neurotechnology in relation to the promotion and protection of all human rights. In February 2023, the Human Rights Council Advisory Committee established a drafting group. In September 2024, a report was submitted containing practice-oriented recommendations on how the Human Rights Council, its special procedures and subsidiary bodies could address the human rights impact of neurotechnology in a coherent, comprehensive and inclusive manner. The Advisory Committee concluded that it is advisable to set out the main principles and applicable standards in a new non-binding document. This document should provide specific guidance including human rights principles and standards related to neurotechnology. The report considered that a new international document from the UN Human Rights Council would be very timely and would be of great value in guiding States and other stakeholders in their policies (Consejo de Derechos Humanos, 2024).

We also note that in the last presentation given by the UN Special Rapporteur on Privacy to the Assembly, she presented a proposal to update UN Resolution 45/95 of 1990, on principles for the processing of personal data. This takes into account, among other things, the new challenges presented by the development of emerging technologies, including neurotechnologies and their convergence with artificial intelligence, in the collection of personal data, among other aspects (Nougrères, 2024). Also, recently (March, 2025) in a report to the 58th session of the Human Rights Council, UN Special Rapporteur calls for regulation of neurotechnologies to protect right to privacy.

On the other hand, we also mentioned that specifically in relation to the childhood approach, research is being developed on: What impact does neurotechnology have on today's children and what could it mean for them in the future? (UNICEF, 2024). A more advanced report is expected to be published in the coming months.

The international frameworks that have been developed recently have prompted States to begin to develop regulatory changes to ensure guarantees for the protection of the most basic human rights of the population. It is for this reason that we are interested in delving into the different strategies that have been adopted by the States to regulate this matter. We will focus on the scenario in the Americas, highlighting that it has been working at the forefront of regulatory initiatives in this regard.

1. Different strategies implemented in the Americas

As a starting point, we clarify that this brief article does not intend to be a critical study of the methodologies considered conceptually ideal for implementing this type of initiative. We will simply develop the different ways in which the sectors in charge of generating public policies are implementing certain measures to regulate neurotechnologies.

We found that the strategies through which progress is being made can be grouped into the following: a) Constitutional Reforms, b) Comprehensive Bills on Neuroprotection, c) Draft Reforms of Personal Data Laws, d) Draft Reforms of Consumer Laws, e) Administrative Regulatory Reforms, f) Other lines of proposals.

1.1. Constitutional reforms

One of the strategies used to implement regulatory changes to expand rights has been through constitutional reforms. The reforms that have already entered into force are those of Chile and the State of Rio Grande do Sul (Brazil). In addition, constitutional reform projects have been presented in Brazil and Mexico.

a) Chile: in Chile, in October 2019, at the initiative of the Future Commission of the Senate and the commitment of the President of the Republic, a constitutional reform bill (Bulletin 13827-19) was introduced for legislative debate. Two years later, the constitutional reform was approved and published in the official gazette on October 25, 2021 (Chile, Promulgación: 14-OCT-2021. Publicación: 25-OCT-2021., 2021). The reform added a new clause to Article 19 Nº1 stating that: "Scientific and technological development will be at the service of people and will be carried out with respect for life and physical and psychological integrity. The law shall regulate

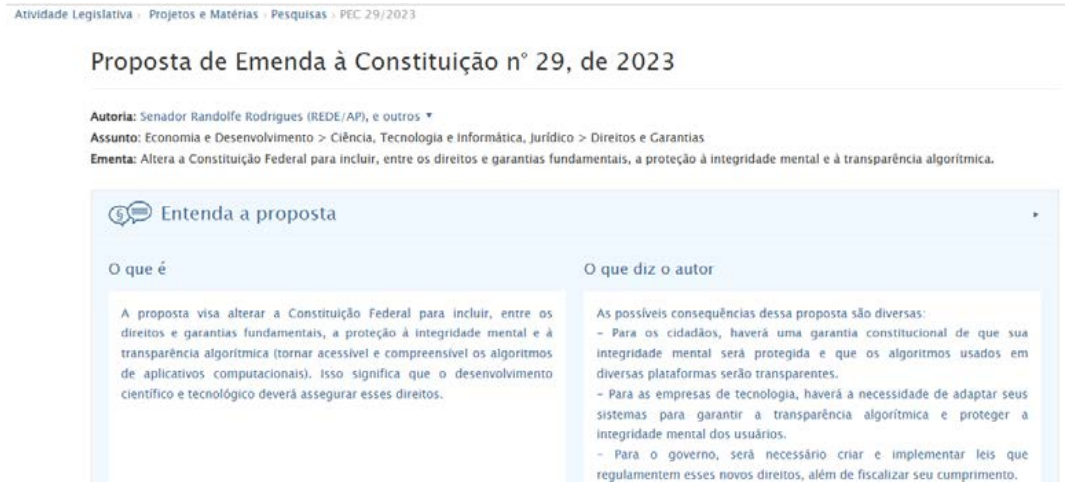
the requirements, conditions and restrictions for its use on people, and shall especially protect brain activity, as well as the information derived from it".

b) State of Rio Grande do Sul (Brazil): in Brazil, on December 20, 2023, the State of Rio Grande do Sul incorporated neuro-rights into its constitution. The reform states that every individual has the right to mental integrity, as an inalienable protection against manipulation resulting from advances in neuroscience and neurotechnology. The reform was approved unanimously (PEC 298 of 2023) and modified the sole paragraph of article 235 of the State Constitution, which deals with science and technological research policy, establishing the protection of mental identity against any research that affects the brain and its activity, without the consent of the individual.

The reform was driven by the commitment of the president of the Legislative Assembly, Deputy Vilmar Zanchin, who led this project in the legislative body.

c) Brazil: Likewise, in Brazil in June 2023, a constitutional reform project (Proposta de Emenda à Constituição n° 29, de 2023) was presented by Senator Randolfe Rodrígues, leader of the Government in the Brazilian National Congress, to incorporate neuro-rights into the Brazilian constitution (Brasil, Proposta de Emenda à Constituição n° 29, de 2023, 2023).

This proposal states that "scientific and technological development will ensure mental integrity and algorithmic transparency, under the terms of the law".



d) México**: i**n Mexico, on August 1, 2023, a draft amendment to Article 4 of the Political Constitution of the United Mexican States was presented that seeks to protect the mental integrity of individuals. The initiative was led by Deputy María Eugenia Hernández Pérez (México, Propuesta de Reforma Constitucional, 2023).

The project proposes the incorporation of the following text: "Every person has the right to full and integral individual identity, as well as to physical and psychological integrity as conditions of his freedom. The state shall guarantee respect for the privacy and mental integrity of persons. No authority or individual may, through the use of any technological mechanism, modify, reduce or affect such integrity and identity."

As a precedent of the neuro-rights movement in Mexico, progress had already been made on the subject with the drafting of the "Charter of Rights of the Person in the Digital Environment" of the Coordination of Personal Data of the National System of Transparency, Access to Public Information and Protection of Personal Data. Chapter VII of this charter incorporates neuro-rights, establishing guidelines for the preservation of digital identity, privacy of neural data, non-interference of neurotechnologies in personal autonomy and freedom of decision. It also includes the principle of equity in access to brain enhancement (SNT, 2023).

CÁMARA DE DIPUTADOS
LXV LEGISLATURA

MARÍA EUGENIA HERNÁNDEZ PÉREZ
DIPUTADA FEDERAL
LXV La Legislatura de la Paridad, la Inclusión y la Diversidad
"2023, año de Francisco Villa, el revolucionario del pueblo"

INICIATIVA CON PROYECTO DE DECRETO POR LA QUE SE ADICIONA UN NOVENO PARRAFO Y SE RECORREN LOS SUBSECUENTES DEL ARTÍCULO 4°; DE LA CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS

La que suscribe, Diputada María Eugenia Hernández Pérez, integrante del Grupo Parlamentario de Morena en la Cámara de Diputados del H. Congreso de la Unión, de conformidad con lo establecido en los artículos 71, fracción II y 78 de la Constitución Política de los Estados Unidos Mexicanos; 55 fracción II, 57 y 63 del Reglamento para el Gobierno Interior del Congreso General de los Estados Unidos Mexicanos, someto a la consideración de esta Honorable Asamblea la siguiente INICIATIVA CON PROYECTO DE DECRETO POR LA QUE SE ADICIONA UN NOVENO PARRAFO Y SE RECORREN LOS SUBSECUENTES DEL ARTÍCULO 4°; DE LA CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS, en materia de Neuroderechos, con base en la siguiente:

1.2. Comprehensive draft laws on Neuroprotection

Another strategy used to regulate these issues is through comprehensive legislation on neuroprotection. We have not yet registered the approval of any law with these characteristics. However, there are projects in the Americas that are similar in proposing a general framework of legislative protection. Those that we were able to register are the draft laws in Chile, Colombia, Ecuador and Mexico. Work is also being done on the drafting of a comprehensive law in the State of Colorado, in the United States.

a) Chile: in Chile, also in 2019, a more detailed bill (Boletín 13828-19) was presented in order to satisfactorily regulate the content of the constitutional reform that enshrines the protection of mental integrity and indemnity in relation to the advancement of neurotechnologies (Chile, Proyecto de Ley Protección de los neuroderechos, 2019).

The project understands that, among other aspects, "neuro-rights" encompass mental privacy and the right to identity, while maintaining the individuality of persons, and defines them as follows: "Article 2 [...] Neurorights: New human rights that protect the privacy and mental and psychic integrity, both conscious and unconscious, of people from the abusive use of neurotechnologies [...]"

This bill, since its entry into Congress, has followed an extensive procedure that continues its course, but with difficulties in moving forward (Chile, Trámite en Senado de la República de Chile, 2024).

12 / 369	05/04/2021	Cuenta del Mensaje 034-369 que hace presente la urgencia Simple	Primer trámite constitucional / Senado
	08/04/2021	Boletín de indicaciones .	Primer trámite constitucional / Senado
25 / 369	04/05/2021	Cuenta del Mensaje 072-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
36 / 369	01/06/2021	Cuenta del Mensaje 104-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
52 / 369	06/07/2021	Cuenta del Mensaje 154-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
60 / 369	03/08/2021	Cuenta del Mensaje 182-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
60 / 369	03/08/2021	La Sala acuerda un nuevo plazo para presentar indicaciones hasta el 4/08/2021, a las 12:00 horas, en la Secretaría de la Comisión.	Primer trámite constitucional / Senado
	03/08/2021	Nuevo plazo para indicaciones. Se amplía plazo para presentar indicaciones hasta el 09/08/2021 a las 12:00 en la Secretaría de la Comisión de Desafíos del Futuro, Ciencia, Tecnología e Innovación	Primer trámite constitucional / Senado
71 / 369	31/08/2021	Cuenta del Mensaje 216-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
79 / 369	28/09/2021	Cuenta del Mensaje 242-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
87 / 369	02/11/2021	Cuenta del Mensaje 270-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
100 / 369	30/11/2021	Cuenta del Mensaje 302-369 que retira y hace presente la urgencia Simple	Primer trámite constitucional / Senado
	06/12/2021	Segundo informe de comisión de Desafíos del Futuro, Ciencia, Tecnología e Innovación.	Primer trámite constitucional / Senado
103 / 369	07/12/2021	Cuenta de segundo informe de comisión .	Primer trámite constitucional / Senado
103 / 369	07/12/2021	Discusión particular . Aprobado	Primer trámite constitucional / Senado
	07/12/2021	Oficio de ley a Cámara Revisora .	Primer trámite constitucional / Senado
109 / 369	13/12/2021	Cuenta de proyecto . Pasa a Comisión de Futuro, Ciencias, Tecnología, Conocimiento e Innovación.	Segundo trámite constitucional / C.Diputados
109 / 369	13/12/2021	Cuenta del Mensaje 391-369 que retira y hace presente la urgencia Suma	Segundo trámite constitucional / C.Diputados
116 / 369	03/01/2022	Cuenta del Mensaje 421-369 que hace presente la urgencia Simple	Segundo trámite constitucional / C.Diputados
126 / 369	26/01/2022	Cuenta del Mensaje 477-369 que retira y hace presente la urgencia Discusión inmediata	Segundo trámite constitucional / C.Diputados
128 / 369	01/03/2022	Cuenta del Mensaje 491-369 que hace presente la urgencia Discusión inmediata	Segundo trámite constitucional / C.Diputados
80 / 372	30/09/2024	Cuenta de Oficio de la Comisión de Futuro, Ciencias, Tecnología, Conocimiento e Innovación (N° 100/19/2024), por el cual solicita recabar el asentimiento de la Sala para que el proyecto sea radicado en la Comisión de Derechos Humanos y Pueblos Originarios, por cuanto consagra los neuroderechos como nuevos derechos fundamentales. RECHAZADO.	Segundo trámite constitucional / C.Diputados
84 / 372	07/10/2024	Cuenta de Oficio de la Comisión de Futuro, Ciencias, Tecnología, Conocimiento e Innovación (N° 101/19/2024), por el cual solicita recabar nuevamente el acuerdo de la Sala para que el proyecto sea radicado en la Comisión de Derechos Humanos y Pueblos Originarios, teniendo en especial consideración que consagra los neuroderechos como nuevos derechos fundamentales. RECHAZADO.	Segundo trámite constitucional / C.Diputados
	07/10/2024	Oficio N° 19919. La Sala rechazasolicitud de la Comisión de Futuro para remitir el proyecto de lev a la Comisión de Derechos Humanos.	Segundo trámite constitucional / C.Diputados

b) Colombia: in Colombia, since mid-2024, a working group has been preparing a bill to be submitted to Congress. It is well advanced and has received contributions from various experts. The bill regulates the principles of neurosciences, neurotechnologies and human rights**.**

It is interesting to mention that within the grounds expressed in the bill, it is imperative to foresee the possible risks in the use of neurotechnologies. Therefore, it is necessary to identify the constitutional provisions and the fundamental rights and public liberties that may be affected. However, it is clarified that "it is not obvious that this interference can be explicitly contemplated in specific rights that have found a place in the Constitution" (Romeo Casabona, 2020). Hence, it is necessary either to reinterpret the norms concerning fundamental rights in accordance with the Universal Declaration of Human Rights and ratified international treaties and agreements or to recognize new rights generically referred to as emerging rights (Colombia, Proyecto

de ley mediante el cual se establecen principios en materia de neurociencias, neurotecnologías y derechos humanos, 2024).

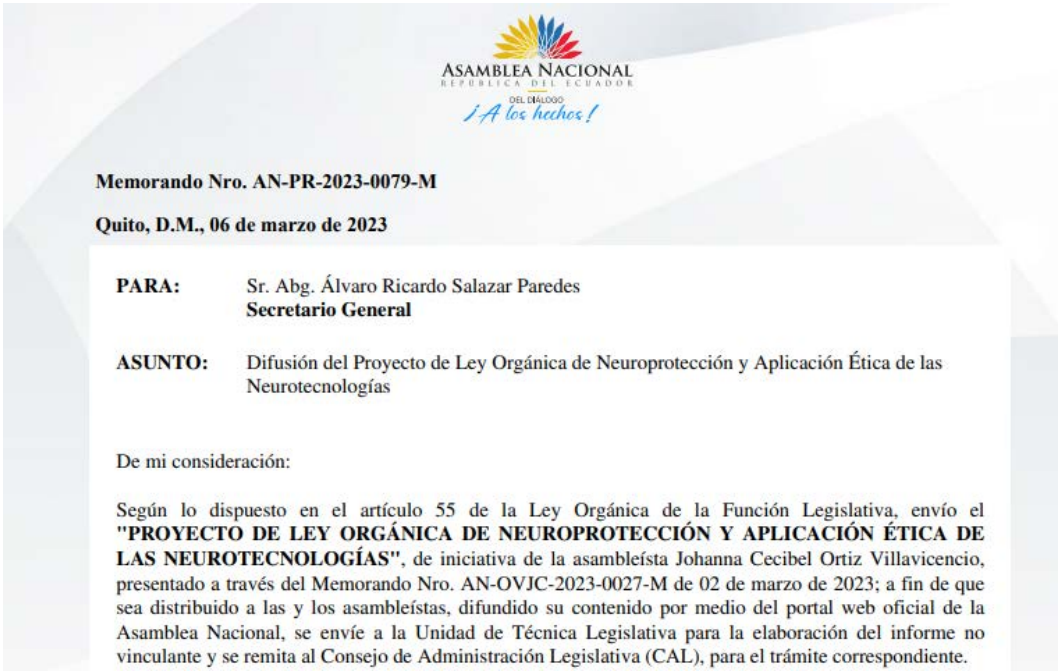
In the same sense, the justification of the project continues that since the seventies, through resolution 3384 of 1975, the United Nations (UN) has recognized that "scientific and technological achievements may pose a danger to the civil and political rights of the individual or group and to human dignity". Therefore, according to the doctrine, it is imperative to adopt measures to avoid the possible negative consequences of certain technological developments for society in general, human rights and human dignity (Remolina, 2024).

Regarding the proposed text of the law, as a starting point, it is suggested that on the one hand, it closely follows (with only minor modifications) the Declaration of Inter-American Principles on Neurosciences, Neurotechnologies and Human Rights. On the other hand, it incorporates other principles that are part of other *softlaw* documents that refer to research and use of biomedical technologies.

c) Ecuador: in Ecuador, in March 2023, a bill was presented (Ecuador, 2023), which according to its own rationale:

> [...] must be understood from the law in its integrality, considering the norms, principles and values that seek, in a transversal manner, to provide an effective response to the protection of these new fundamental rights that are born in the light of the progress of sciences and technologies and whose results and consequences can already be seen in the field of medicine. To this end, this bill has a strong anchorage in human dignity as an underlying principle to which neurotechnology must always look, incorporating, in addition, an element of equal access to the development of the technique, which is materialized through equal access to the increase of mental capacity, to avoid any hint of arbitrary and unlawful differentiations.

The bill states that advances in neuroscience and neurotechnology pose ethical and regulatory challenges to the fundamental rights of individuals and could even rethink what it means to be human. However, such technologies lack an adequate legal framework to protect fundamental rights from the challenges posed by these advances (Ecuador, 2023).



ASAMBLEA NACIONAL
REPÚBLICA DEL ECUADOR
DEL DIÁLOGO
¡A los hechos!

Memorando Nro. AN-PR-2023-0079-M

Quito, D.M., 06 de marzo de 2023

PARA: Sr. Abg. Álvaro Ricardo Salazar Paredes
Secretario General

ASUNTO: Difusión del Proyecto de Ley Orgánica de Neuroprotección y Aplicación Ética de las Neurotecnologías

De mi consideración:

Según lo dispuesto en el artículo 55 de la Ley Orgánica de la Función Legislativa, envío el **"PROYECTO DE LEY ORGÁNICA DE NEUROPROTECCIÓN Y APLICACIÓN ÉTICA DE LAS NEUROTECNOLOGÍAS"**, de iniciativa de la asambleísta Johanna Cecibel Ortiz Villavicencio, presentado a través del Memorando Nro. AN-OVJC-2023-0027-M de 02 de marzo de 2023; a fin de que sea distribuido a las y los asambleístas, difundido su contenido por medio del portal web oficial de la Asamblea Nacional, se envíe a la Unidad de Técnica Legislativa para la elaboración del informe no vinculante y se remita al Consejo de Administración Legislativa (CAL), para el trámite correspondiente.

d) México: in Mexico, a comprehensive project on neuroprotection was also presented in September 2023 (México, Proyecto de Ley General de Neuroderechos, 2023) . It proposes the creation of a General Law on Neurorights and Neurotechnologies that establishes principles and objectives to protect human dignity and rights related to the nervous system and brain activity.

The bill was introduced by the Senate at the proposal of Senator Alejandra Lagunes Soto Ruíz. The initiative expresses several provisions on neuro-rights and neuro-technologies and aims to propose the bases and general principles to guarantee the protection of human dignity and the recognition, respect, promotion and defense of human rights related to the central and peripheral nervous system.

The proposal had the support of the National Institute for Transparency, Access to Information and Protection of Personal Data (INAI) as Mexico's guarantor body for privacy and personal data protection.

1.3. Draft amendments to Personal Data Laws

On the other hand, progress was also made in the design of legislative strategies proposing the reform of laws that protect personal data. There are bills in this regard in Brazil, Colombia and Uruguay.

a) Brazil: in Brazil, this bill proposes the amendment of Law No. 13,709 of August 14, 2018 (General Law on Personal Data Protection), in order to conceptualize neural data and regulate its protection (Brasil, Cámara de Diputados, 2022).

The project highlights in one of its proposals the incorporation of neural data as sensitive data, stating that: "II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biometrico, dado neural, quando vinculado a uma pessoa natural;" (Brasil, Cámara de Diputados, 2022) .

b) Colombia: in Colombia, in August 2024, a draft law on personal data was presented (Colombia, Cámara de Representantes, 2024).

It incorporates several references to neurodata, including:

> Article 4°. Definitions. For the purposes of this law, the following definitions shall apply: [...] 12. "Sensitive data": are those that affect the privacy of the holder or whose improper use may generate discrimination, such as those that reveal racial or ethnic origin, political opinions, religious or philosophical convictions and union membership, social organizations, human rights, genetic data, neurodata [...];
>
> 18. "Neurodata": for the purposes of this law, refers to the set of information obtained from the brain activity and the nervous system, as well as the information inferred from this data, of an identified or identifiable natural person.
>
> "Article 10. Processing of sensitive data. The processing of personal data revealing .... genetic data, neurodata, biometric data aimed at univocally identifying a natural person, data relating to health, data relating to sex or biological characteristics, gender identity or expression and sexual orientation of a natural person is prohibited".
>
> "The data protection impact assessment referred to in paragraph 1 shall be required in particular when: [...] d) Processing of neurodata is carried out".
>
> "CHAPTER II Neurodata Article 56. Processing of neurodata. The controller, or where appropriate the processor processing neurodata, shall ensure the protection of fundamental rights and guarantees related to the protection of personal data and privacy of individuals [...]"

In addition, recently and with wording similar to the draft law previously mentioned, in the Senate was presented the draft law No. 395, on March 11, 2025, by which the principles in the field of neurosciences, neurotechnologies and human rights are regulated.

c) Uruguay: in Uruguay, in June 2024, a reform project was presented to incorporate neurodata as sensitive data. The project under folder number 4393 proposes under a single article the following: **"**Section E) of article 4 of Law No. 18.331, of August 11, 2008, is hereby amended to read as follows: 'Sensitive data: personal data revealing racial and ethnic origin, political references, religious or moral convictions, union membership and information concerning health or sexual life, brain and neuronal activity of persons'". Montevideo, June 10, 2024, Rodrigo Goñi Reyes, Representative for Montevideo (Uruguay, 2024).

1.3. Consumer Law Reform Bills

In addition, there are other strategies by which legal reforms are being implemented to protect neurodata. This path is being developed through reforms of laws that protect consumer rights. In fact, in the States of Colorado and California, reforms in this sense have been approved (Foundation Neurorights in the US, 2024).

The UN Human Rights Council Report states that "Countries that adopt a more market-oriented approach, such as the United States, tend to focus on technical issues to ensure the safety and security of technological developments, while emphasizing consumer rights" (Consejo de Derechos Humanos, 2024).

a) Colorado: In the State of Colorado, in the United States, on April 17, 2024, a legislative reform to the State's Consumer Protection Law was approved (Colorado, 2024).

By way of reform notice that, in 2021, the general assembly enacted Senate Bill 21-190, which established the "Colorado Privacy Act" as part of the "Colorado Consumer Protection Act"; and that the "Colorado Privacy Act" protects the privacy of consumers by establishing specific requirements for entities that process consumers' personal data.

It also clarifies that the "Colorado Privacy Act" establishes certain rights that consumers may exercise with respect to the processing of their personal data and includes greater protections for data collected about bodily or mental functions.

Therefore, the general assembly (Colorado, 2024) determined that it was necessary and appropriate to expand the "Colorado Privacy Act" definition of "confidential data" to include: [...] (b) "Neural data", which is information generated by measuring the activity of an individual's central or peripheral nervous systems and which may be processed by or with the assistance of a device.

b) California: The State of California was the second jurisdiction in the United States, after Colorado, to define and protect consumer neural data as sensitive personal information.

SB 1223 added "neural data," i.e., "information generated by measuring the activity of a consumer's central or peripheral nervous system, and not inferred from non-neural information," as a category of sensitive personal information for purposes of the California Consumer Privacy Act (CCPA), giving this information additional protections under the law (California, 2024): "This bill would define "sensitive personal information," for purposes of the CCPA, to additionally include a consumer's neural data, and would define "neural data" to mean information that is generated by measuring the activity of a consumer's central or peripheral nervous system, and that is not inferred from nonneural information."

In February 2024, California State Senator Josh Becker introduced Senate Bill 1223 and in May, it passed the California State Senate by a unanimous vote of 38-0 and passed the State Assembly, which referred it to the State Assembly Committee on Privacy and Consumer Protection. In that body, it passed unanimously (10-0). Finally, on September 28, 2024, the law establishing very strong protections for neural data was enacted.

1.4. Administrative regulatory reforms

Other strategies include regulating the issue through administrative regulations. This is the case in Peru, through the Ministry of Justice and Human Rights.

a) Perú: in Peru, within the framework of the participatory process opened by Ministerial Resolution No. 0270-2023-JUS of August 25, 2023, the draft Regulations of Law No. 29733, Personal Data Protection Law, and its supporting Explanatory Memorandum were published. In said regulation, mental health data and emotional data were incorporated as sensitive data, allowing them to have greater guarantees and legal protections (Perú, Resolución Ministerial N° 0270-2023-JUS, 2023).

Following the open public consultation process for the drafting of the Regulations, in November 2024, through Supreme Decree 016-2024-JUS, the new Regulations were approved, which provides (Perú, 2024):

> "[...] The purpose of these Regulations is to establish provisions for the proper application of Law No. 29733, Personal Data Protection Law, hereinafter the Law, in order to guarantee the fundamental right to the protection of personal data, regulating an appropriate treatment by natural persons, public entities and institutions belonging to the private sector, particularly in the digital environment [...]"
>
> "[...] 6. Sensitive data: Information relating to genetic or biometric data of the natural person, neuronal data, moral or emotional data, facts or circumstances of their affective or family life, personal habits that correspond to the most intimate sphere, information relating to union membership, physical or mental health or other similar that affect their privacy [...]"

1.5. Other lines of proposals

To conclude with the different ways chosen to regulate, we mention that there were also some initiatives in Argentina.

One of them in the area of criminal procedural reform, bill 0339-D-2022, introduced in the Chamber of Deputies of Argentina, proposes a series of amendments regarding the inclusion of brain imaging techniques and other neurotechnologies as evidence in criminal proceedings, provided that these techniques allow inferring the mental activity of a person based on his brain structure or function. These evidentiary means may only be used in the process by court order and with the consent of the person concerned. The bill also provides for the amendment of Law 24.660, relating to imprisonment. In this case, the proposed amendment refers to social rehabilitation treatments for inmates, which, when they include neurotechnologies, must again have a court order and the explicit consent of the inmate. This modification also appeals to the omission of cognitive and algorithmic biases.

In addition, a proposal was presented in the Senate (S-2446/2023) for the study of these issues through a specific Legislative Commission. The project proposes the creation of a Bicameral Commission for the regulatory treatment of the comprehensive protection of neuro-rights, since this is a matter on which there is no specific regulatory framework, and which poses legal and ethical challenges (Argentina, 2023).

On the other hand, progress has been made in legislative reforms in the States of Illinois, Massachusetts and Montana in the USA.

2. Is it being regulated in other continents?

In other continents as well, for some time now, different initiatives have been developed that propose to provide a framework of protection in this area.

It is noted that, in **Europe**, the Council of Europe has initiated a reflection on the subject and may adopt measures based on its strategic action plan on biomedical technologies. The European Union has not proposed any policy specifically adapted to neurotechnologies, but has proclaimed, in the Leon Declaration on European Neurotechnology (of October 2023), the need for a human-centered and rights-oriented approach (Consejo de Derechos Humanos, 2024) .

It is interesting to note the creation of the Charter of Digital Rights, in 2020 in Spain, where a participatory process was developed for its elaboration, particularly in section XXIV, the Charter addresses digital rights in the use of neurotechnologies. The Charter of Digital Rights warns of the need for a specific regulation that establishes conditions, limits and guarantees in the implementation of neurotechnologies. On the other hand, it establishes to guarantee, among other rights, identity, self-determination, integrity, equality and non-discrimination, and to ensure the confidentiality and security of data (Carta de Derechos Digitales, España, 2020).

Also in Spain, in December 2023, the Office of Science and Technology of the Spanish Chamber of Deputies published a Summary C: Advances in neuroscience: applications and ethical implications. The disruption of neurotechnologies. Summary C: Advances in neuroscience: applications and ethical implications We cite it because it refers to the work being carried out at the regional level, although there are some inaccuracies in certain projects mentioned (Avances en neurociencia: aplicaciones e implicaciones éticas, 2023).

Similarly, in the UK, in 2022, the Information Commissioner's Office report entitled "Technological Futures", highlighted certain issues that broadly converge on common points to ensure that safety and security standards are met in neurotechnology developments (UK, 2022).

In France, a bioethics-related law complements the public health law to prohibit any activities related to the modification of brain activity that present - or may be suspected of presenting - a serious danger to human health. France has adopted a law relating to bioethics and has enacted a charter of voluntary measures for the responsible development of neurotechnologies (Consejo de Derechos Humanos, 2024).

On the other hand, it highlighted the UN Human Rights Council working group that **China** adopted a set of specific guidelines on research ethics in the context of the brain-computer interface (Consejo de Derechos Humanos, 2024).

3. By way of conclusion

After analyzing the different initiatives that are being presented with the aim of protecting and safeguarding the rights of the population. We are interested in highlighting ten (10) points that we consider relevant when thinking about any of the strategies for the design of public policies for the protection of human rights, in view of the advance of the development of neurotechnologies.

a) **The information from our central and peripheral nervous system is invaluable**.

Neurotechnologies do not register what we think, they only detect patterns that generate information whose value and importance we have not yet been able to mea-

sure. That is why there is a need to design a new paradigm in convergence with Artificial Intelligence.

b) **Insufficient self-regulation and self-control**.

At present, with the development of neurotechnologies, situations arise that must be regulated exclusively through public policies, since they escape from everything known, they are unprecedented. The regulation of relations with users cannot be left in the hands of consumer companies alone.

c) **Neurotechnology for "medical use" vs. neurotechnology for "consumption"**.

The weakness in rights protection systems lies in consumer neurotechnologies. This is where guarantees must be reinforced and public policies must be prioritized.

d) **Protection of human rights by design**.

As stated in Principle 2 of the OAS Inter-American Juridical Committee, comprehensive protection and respect for human rights must be ensured from design to implementation, marketing and use (patients, users and/or consumers).

e) **Neural data as sensitive personal data**.

We believe that the existence of this new category of protection is substantial. We understand that it should be an independent category within the protection of sensitive personal data, because beyond the identification of the person (biometric data), neurodata have their own characteristics and need to be well defined by the regulation.

f) **Express and informed consent**.

The consent of the person who is the owner of the neurodata is an essential requirement for access to the collection of brain information; it must be free, informed, express, unequivocal, free of vices and specific. That is to say, the consent must take into account the whole path and the destinations that will be given to the neurodata, it is observed that there are cases where a single generic consent is given without giving many precisions about the course of the neurodata, that is why each of the specificities in the use of the data must be clarified and, in that case, separate consents must be given for each one. In addition to always having the possibility to revoke it.

g) **The importance of a supervisory and control authority**.

We believe it is necessary to point out that any regulation in this area must have a control authority in charge of monitoring and supervising compliance with the standard and supervising the operation of neurotechnology developments. In addition to intervene in case of irregularities. The authority should take into account the strategy of each country, they can refer to existing authorities that take on a new competence or to be formed

h) **Compliance with human rights and business guiding principles**.

In accordance with the Inter-American Standards on Business (REDESCA, 2019) and Human Rights and the Guiding Principles on Business and Human Rights (ONU, 2011), States should establish a national framework, including oversight mechanisms, and exercise their due diligence obligations with respect to neurotechnology companies (ONU, 2011). Likewise, business enterprises should develop effective tools and seek appropriate advice to integrate a human rights-based approach

and undertake risk assessments on actual and potential human rights impacts, both direct and indirect, during all phases of their operations (Consejo de Derechos Humanos, 2024).

i) **Protecting democratic systems**.

There is a need to protect democratic systems in the face of new dilemmas imposed by the scientific advances of neurotechnologies and their convergence with AI and immersive technologies, we must balance innovation with the protection of citizens' rights, because it directly impacts the functioning of democracy.

j) **Vulnerable groups**.

It is necessary to protect identity and agency (the latter understood from its sociological meaning: the ability to choose our actions with free will). On the other hand, there is concern about providing greater protection to vulnerable groups, due to various factors, including the possibility of "artificial augmentation" of brain capacities (which could produce inequities). On this same argument, raises the need to establish clear limits and exercise greater control over the augmentation of cognitive capacities. In this sense, the protection of the right to equality provides for equitable access to neurotechnologies and guarantees non-discrimination of categories that have historically been subject to discrimination: race, color, gender, nationality, religion, social status, among others.

References

Argentina. (2023). Obtenido de Comisión Bicameral Neuroderechos: https://www.senado.gob.ar/parlamentario/DAEDIGITAL/abrirVerPdf/F1936771702920571_S2446.pdf

Brasil. (2022). *Cámara de Diputados*. (P. 522/2022, Editor) Obtenido de https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2317524&fichaAmigavel=nao

Brasil. (2023). Proposta de Emenda à Constituição n° 29, de 2023. Obtenido de https://www25.senado.leg.br/web/atividade/materias/-/materia/158095

California. (2024). *California Privacy Protection Agency*. Obtenido de https://files.lbr.cloud/public/2024-09/20230SB1223_95.pdf: https://cppa.ca.gov/meetings/materials/20240716_item7_sb_1223.pdf

Carta, D. (2020). *Carta de Derechos Digitales, España*. Obtenido de https://espanadigital.gob.es/lineas-de-actuacion/carta-de-derechos-digitales

Chile. (2019). Proyecto de Ley Protección de los neuroderechos. Obtenido de https://www.diarioconstitucional.cl/wp-content/uploads/2020/12/boletin-13828-19-nuroderechos.pdf

Chile. (2021). *Promulgación: 14-OCT-2021. Publicación: 25-OCT-2021.* Materias: Desarrollo Científico, Desarrollo Tecnológico, Actividad Cerebral, Respeto a la Vida, Respeto a la Integridad Física, Respeto a la Integridad Psíquica. Obtenido de https://www.bcn.cl/leychile/navegar?idNorma=1166983&idParte=10278855&idVersion=2021-10-25

Chile. (2024). *Trámite en Senado de la República de Chile*. Obtenido de https://tramitacion.senado.cl/appsenado/templates/tramitacion/index.php?boletin_ini=13828-19

Colombia. (2024). *Cámara de Representantes*. (152/2024C, Editor) Obtenido de https://www.camara.gov.co/regimen-proteccion-de-datos-personales

Colombia. (2024). Proyecto de ley mediante el cual se establecen principios en materia de neurociencias, neurotecnologías y derechos humanos. *Notas preliminares para continuar construyendo un proyecto de ley Ana Isabel Gómez Córdoba, Diana Rocío Bernal Camargo Senador Carlos Julio González y Nelson Remolina Angarita.*

Colorado. (2024). *HOUSE BILL 24-1058*. Obtenido de https://leg.colorado.gov/sites/default/files/2024a_1058_signed.pdf

Comité Jurídico Interamericano, O. (2023). *Principios Interamericanos Neurociencias, Neurotecnologías y Derechos Humanos.* Obtenido de https://www.oas.org/es/sla/cji/docs/CJI-RES_281_CII-O-23_corr1_ESP.pdf

Consejo de Derechos Humanos, O. (2024). *Efectos, oportunidades y retos de la neurotecnología*. Informe del Comité Asesor del Consejo de Derechos Humanos. Obtenido de https://docs.un.org/es/A/HRC/57/61

Diputados. (2023). *Avances en neurociencia: aplicaciones e implicaciones éticas*. Obtenido de https://oficinac.es/es/informes-c/neurociencia-aplicaciones

Ecuador. (2023). *Proyecto de Ley Orgánica de Neuroprotección y aplicación ética de las neurotecnologías*. Obtenido de https://participa.asambleanacional.gob.ec/legislation/processes/64/draft_versions/81

México. (2023). Propuesta de Reforma Constitucional. Iniciativa con Proyecto de Decreto por la que se adiciona un novena párrafo y se recorren los subsecuentes del artículo 4°; de la Constitución Política de los Estados Mexicanos. Obtenido de http://sil.gobernacion.gob.mx/Archivos/Documentos/2023/08/asun_4588906_20230808_1690903141.pdf

México. (2023). *Proyecto de Ley General de Neuroderechos*. Obtenido de https://infosen.senado.gob.mx/sgsp/gaceta/65/3/2023-09-26-1/assets/documentos/Inic_PVEM_diversos_senadores_art_73_CPEUM.pdf

Monti, Natalia; Sánchez, Moisés & Colombara, Ciro. (2023). Neurotecnologías en contextos democráticos. (FLACSO, Ed.) *Revista Estado y Políticas Públicas*, 73-95.

Neuralink. (2024). Obtenido de https://www.youtube.com/watch?v=ZzNHxC96rDE

Nougrères, A. B. (2024). Tercera Comisión, 27ª sesión plenaria, Asamblea General, 79° período de sesiones. Obtenido de https://webtv.un.org/en/asset/k1s/k1sdvaiomu

ONU. (2011). *Principios rectores sobre empresas y los derechos humanos.* HR/PUB/11/04. Obtenido de https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_SP.pdf

Perú. (2023). *Resolución Ministerial N.° 0270-2023-JUS*. Obtenido de https://www.gob.pe/institucion/minjus/normas-legales/4574374-0270-2023-jus

Perú. (2024). *DECRETO SUPREMO N° 016-2024-JUS*. Obtenido de https://img.lpderecho.pe/wp-content/uploads/2024/11/Decreto-Supremo-016-2024-JUS-LPDerecho.pdf

Playing Video Games With Mind Control. (2023). Obtenido de https://youtu.be/DBYY3D1gkQ0

REDESCA. (2019). *Informe Empresas y Derechos Humanos: Estándares Interamericanos*. Comisión Interamericana de Derechos Humanos, Relatoría Especial sobre Derechos Económicos Sociales Culturales y Ambientales. OEA. Obtenido de https://www.oas.org/es/cidh/informes/pdfs/EmpresasDDHH.pdf

Remolina, N. (2024). Neuro reflexión : hacia una Declaración Universal sobre las neurotecnologías y los derechos humanos. En *En defensa de los neuroderechos* (pág. Pág 224). Editado por la Fundación Kamanau.

Romeo Casabona, C. M. (2020). Bioconstitución. *Revista de Derecho y Genoma Humano*: Genética, Biotecnología y Medicina Avanzada,53, 15-21.

SNT. (2023). Carta de Derechos de la Persona en el Entorno Digital. Comisión de Protección de Datos Personales del Sistema Nacional de Transparencia. Obtenido de https://snt.org.mx/cpdp/wp-content/uploads/Carta-DDigitales_NuevoDiseno_FINAL_DIGITAL.pdf

UK. (2022). *ICO tech futures: neurotechnology*. Obtenido de https://ico.org.uk/about-the-ico/research-and-reports/ico-tech-futures-neurotechnology/

UNESCO. (2024). Obtenido de Consulta virtual sobre ética de la neurotecnología en América Latina y el Caribe: https://www.unesco.org/es/articles/consulta-virtual-sobre-etica-de-la-neurotecnologia-en-america-latina-y-el-caribe

UNESCO. (2025). *Ethics of neurotechnology*. Obtenido de https://www.unesco.org/en/ethics-neurotech/recommendation?hub=83294

UNICEF. (2024). *Neurotechnology and Children.* Obtenido de https://www.unicef.org/innocenti/media/8956/file/UNICEF-Innocenti-Neurotechnology-and-Children-2024.pdf

Uruguay. (2024). *Cámara de Representantes*. Obtenido de https://parlamento.gub.uy/documentosyleyes/documentos/repartido/representantes/49/1152/0/PDF

USA. (2024). *Foundation Neurorights in the US*. Obtenido de https://neurorightsfoundation.org/united-states

— IV —

NEUROTECHNOLOGY AND MINORITIES (DES)PROTECTION

— 12 —

Neurotechnological advancements for fairness: detecting and mitigating implicit racial bias in criminal sentencing

JANNA DA NÓBREGA SOUZA
University of Murcia - Spain
Email: janna.nobrega@um.es

A preliminary overview

Impartiality stands as the cardinal principle that underpins the very essence of judicial function. Within this framework, its significance is inextricably linked to the conception of an absolute absence of bias. In other words, bias—whether objective in nature or merely perceived—represents an inherent contradiction to the ideal of impartiality itself (Mack *et al*., 2021). In this regard, the concept is closely associated with a predominant inclination or an intrinsic predisposition within the individual tasked with the decision-making process, thereby jeopardizing the very objectivity that such an endeavor demands (Geyh, 2014).

Within the specific domain of criminal sentencing, this predisposition appears to be circumscribed by a range of potential sources, among which gender and race/ethnicity are particularly prominent (Meyer and Jesilow, 1996). Concerning the former of these factors, empirical evidence consistently demonstrates that women are accorded more favorable treatment in judicial decisions relative to men (Holmes, Feldmeyer, and King, 2020). Scholars substantiate this observation through three theoretical frameworks (limited rationality, focal concerns, and guilt attribution), all of which share a common thread: the relationship between gender and sentencing is regarded as a collateral outcome of flawed human judgment, as evidenced in the judicial results. According to Rodríguez, Curry, and Lee, such errors in judgment stem from temporal constraints that lead to conclusions based on incomplete infor-

mation regarding the accused (Rodríguez, Curry, and Lee, 2006). As a result, women charged with the commission of criminal acts are frequently perceived as posing a lesser threat to the community or a reduced risk of recidivism, whereas men are deemed more culpable and, consequently, more accountable for their transgressions (Steffensmeier *et al*., 1998).

These inter-gender disparities are also evident among individuals of the same sex, with empirical studies revealing that Black men and those from ethnic minority backgrounds are subjected to more severe sentences than their white counterparts (Veiga, Pina-Sánchez, and Lewis, 2022). In the context of England and Wales, for example, it was determined that these accused individuals exhibited a 1.4-fold increased likelihood of being sentenced to a term of imprisonment for drug-related offenses (Uhrig, 2016). The situation appears to be analogous in the United States, where the same individuals are subjected to substantially longer and more severe sentences, particularly within the realm of Federal Courts (Light, 2021). These observations highlight the second of the factors that underpin bias, which will be the focus of analysis in the present study. Broadly speaking, there exists a discernible propensity to associate Afro-descendant or Afro-centric characteristics with traits of danger, negativity, and criminality, a tendency that reflects the outcome of an underlying dynamic associated with implicit racial biases (Du, 2015).

Implicit biases constitute attitudes and stereotypes that involuntarily shape our comprehension, actions, and social conduct (Kang *et al*., 2012). The conceptual framework is predicated upon two fundamental notions. The first pertains to cognitive activation driven by automatic processes of memory and associative links, which are not anchored in conscious, deliberate choices but, rather, in memories and associations that are more readily accessible within our cognitive structure. Such activation transpires beyond the scope of our conscious awareness and may directly contradict the beliefs we consciously espouse. Conversely, the second definition suggests that a particular decision or judgment is influenced by this unconscious dynamic (Saunders and Midgette, 2023). Irrespective of its scope, it is incontrovertible that this phenomenon holds the potential to engender particularly deleterious consequences owing to its inherently automatic nature, thereby functioning independent of the individual's rational volition (Greenwald and Banaji, 1995). Under specific circumstances, this involuntariness is rendered essential for the perception, processing, and retention of information, given that the brain, in its endeavor to categorize the immense volume of data it receives, must optimize the processing of external stimuli (Pyun, 2015). Implicit biases are invariably present, and, as has been previously demonstrated, even during the phase of sentencing imposition. As a result, these biases persist, and as has been previously delineated, even during the imposition of sentences. It is imperative to highlight that neuroscience offers profoundly relevant insights into practical strategies aimed at identifying these subjective influences, providing tools that may prove instrumental in alleviating the negative consequences associated with them. One such approach resides in individualization, which mandates that those endeavoring to counteract their biases meticulously gather specific knowledge concerning the individual before them. Thus, rather than rendering judgments based on group characteristics, a judge, endowed with the insights afforded by neurological data, would be capable of grounding their verdicts in the unique attributes of the individual at hand, thereby transcending the generalizations that limit a just and

objective perspective (Devine *et al*., 2012). In close relation to this practice, the adoption of perspective stands as an additional viable instrument, defined as the capacity to position oneself within the standpoint of an individual belonging to a distinct social or ethical group. In this regard, assuming the perspective of the defendant may facilitate the adjudicator's clearer acknowledgment of the responsibility emanating from their actions, thereby promoting a more profound and nuanced comprehension of the specific circumstances surrounding the case in question (Negowetti, 2015). In light of the foregoing, and prior to embarking upon a comprehensive exposition of the delineated techniques, it is imperative to undertake a meticulous examination of the theories pertaining to the decision-making process, with the objective of, in a subsequent stage, elucidating the tangible impact of neuroscientific advancements on the pursuit of judicial objectivity that is not merely comprehensive and robust, but also rigorously anchored in the most exacting epistemological principles.

1. Judicial decision-making theories: an in-depth analysis of how judges reach veredicts

The theory most broadly endorsed concerning the manner in which judges render their verdicts is that advanced by legal realists, who contend that judges determine the outcome of a judicial dispute even prior to assessing whether the conclusion is rooted in an extant legal principle. More precisely, the adjudicator evaluates the facts presented and, without first undertaking a thorough examination of precedent or applicable statutory provisions, determines how the judgment shall be pronounced. Only upon attaining this outcome does the adjudicator proceed to identify the jurisprudential principles or the pertinent legislative provisions that substantiate and legitimize the resolution rendered (Carpuso, 1998). In view of the foregoing, decisions are not anchored in a systematic appraisal, but rather in a moment of acute discernment, termed judicial intuition, which constitutes the perceptual judgment regarding what is deemed appropriate or inappropriate in a given context (Everson, 1919). The emotional impetus accompanying this phenomenon would, in turn, constitute a response to a series of stimuli originating from the witnesses, which are subsequently filtered through the prejudices and stereotypes intrinsic to their perception, shaped by political movements, psychological biases, or sociological antecedents (Nagel, 1962).

In contrast to this perspective, the formalist theory posits that every judicial opinion is grounded solely in the very existence of the Law. This, whether via precedent or legislative authority, directs the judge's decision, who after subjecting the pertinent facts to a comprehensive analysis, applies the legal provisions with strictness, in an exercise of unwavering adherence to the pre-existing normative framework. This argument is situated within the structure of a deductive logical syllogism: the rule of law constitutes the major premise, the facts the minor premise, and the legal outcome materializes as the conclusion (Huhn, 2003). Thus, the resolution would be uncovered through the application of a mathematical formula, the precision and consistency of which ensure that any other individual, employing the same procedure under analogous circumstances, would arrive at an identical conclusion (Carpuso, 1998). What is posited presupposes that the factual realities emerge from the contradictory testimonies of the parties and are reflected in the independent interpretation of the adjudicator, who, in the course of their duties, seeks a faithful legal represen-

tation of the underlying complexity inherent in the narratives presented (Maroney, 2011). Notwithstanding, although judicial perception may not endure distortions as profound as those that may affect a jury, judges, as intrinsically fallible beings, are not exempt from the possibility of committing errors in their evaluations. Their inherent humanity renders them susceptible to the frailty of issuing imprecise judgments, entangled within the complex labyrinth of intricacies that characterizes the scenario they are called to unravel. Concurrently, the very facts of a case are not fixed, determinable variables, but rather an indeterminate and unpredictable element, the significance of which transcends a definitive verdict without the nuance provided by contextual analysis. These critiques resonate profoundly with the postulates of legal realism, a doctrine that, as delineated earlier, is underpinned by an unyielding commitment to empiricism, accentuating the primacy of the legal process over the inflexibility of prescriptive norms, while concurrently fostering a discerning vigilance against political influences and inherent biases. Moreover, it advocates for a conception of law as a malleable framework, with its ultimate aim residing in the realization of the most elevated and fundamental social objectives (Cohen and Bodansky, 2024). Such principles not only illuminate the intricate dynamics that constitute the judicial decision-making process but also underscore the acute perceptiveness of the legal realists in conceiving the judge as a fully human entity, as opposed to reducing them to a figure abstracted from emotion or essential humanity (Carpuso, 1998).

In this context, it is entirely plausible to establish a connection between this intellectual tradition and cognitive-behavioral studies, conceived as an interdisciplinary paradigm that probes the intricate recesses of the human mind, meticulously examining the diverse processes involved in the acquisition, processing, and storage of information, as well as the complex mechanisms underlying evaluation and decision-making. In this regard, the reflections put forth by Cohen and Bodansky contend that the interdisciplinary relationship with other scientific fields can be realized in two distinct manners, contingent upon whether a legal theory addresses issues analogous to those posited by cognitive and behavioral sciences, or when it raises questions of a fundamentally different nature (Cohen and Bodansky, 2024). Primarily, in a manner akin to the aforementioned studies, certain legal perspectives—such as the theory of rational choice and constructivism—are principally concerned with explanatory issues to which empirical evidence holds significant relevance. The objective of these doctrines is to examine the causal relationships and effects that underpin beliefs and behaviors, while simultaneously seeking to clarify the mechanisms and rationale through which individuals engage with legal arguments. Within this intellectual framework, cognitive-behavioral research emerges as an indispensable tool, endowed with the capacity to both substantiate and critically examine its tenets through the application of rigorous empirical observation. In contrast, other theoretical approaches, such as positivism and natural law, focus on questions of a theoretical and normative nature that pertain to the very essence of the discipline and its prospective future direction. To achieve this, they utilize analytical and interpretive methodologies in an effort to provide insight into these fundamental inquiries. From this standpoint, by formulating disparate inquiries and adopting divergent methodological frameworks, each discipline enhances the other from an external vantage point. In other terms, legal theory is endowed with the capacity to identify phenomena of substantial significance that cognitive science has yet to investigate in depth, which could subse-

quently undergo rigorous examination to ascertain whether they bear a relationship to observable cognitive processes: the internal perspective, the logic of consequences as opposed to the logic of appropriateness, as well as the role of sanctions in contrast to rewards, in terms of their capacity to influence behavioral outcomes.

In contrast, cognitive science has the potential to propel these scholars toward a process of refinement and recalibration of their theoretical frameworks, thereby facilitating the emergence of novel and intellectually enriching avenues for contemplation. With particular regard to its affiliation with legal realism, the pragmatic and pluralistic nature of cognitive science positions it as a methodological approach of exceptional receptivity, poised to seamlessly integrate its empirical insights into the analysis of the processes through which legal norms are constructed, ascribed meaning, and exert tangible effects within the very fabric of society (Cohen and Bodansky, 2024). According to Paterson and Pardo, this heightened focus on socio-scientific inquiry currently occupies a position of considerable privilege, providing an ideal platform for a significant advancement through the integration of neuroscientific breakthroughs, which illuminate the intricate workings of the rational psyche, unveiling its complex dynamics with unequivocal clarity (Paterson and Pardo, 2010).

2. The impact of progress in Neuroscience: reconciving legal cognition

The behavior typically conceived as decision-making entails a complex interplay of cognitive functions that occur deep within the brain, and neuroscience, through its advancements and discoveries, has provided compelling insights into these mechanisms (Yoder and Decety, 2018; Garrigan, Adlam, and Langdon, 2016). Spain Bradley meticulously synthesizes these insights in *The Disruptive Neuroscience of Judicial Choice*. First, it underscores the absence of a singular, unified process, contending that, when the brain engages in the act of decision-making, it mobilizes a multiplicity of regions and circuits that function both sequentially and concurrently, intricately interwoven to form a complex cognitive structure, which ultimately culminates in the rendering of judgment and the subsequent choice. Furthermore, it posits that decision-making is fundamentally organized around the explicit aim of attaining a predefined objective. From a neuroscientific standpoint, this activity is conceptualized as goal-directed thinking, a mechanism wherein, through active engagement, the prefrontal cortex (PFC henceforth) engages. This region, distinguished by its singular sophistication, manifests its capacity to direct attention toward pertinent stimuli, with the resultant process being identified as cognitive control (Spain Bradley, 2018).

A third and final observation pertains to the intricate dynamics through which the mind engages with creative judicial reasoning (Spain Bradley, 2018). Within this framework, Elsbach and Barr contend that positive emotional states are particularly conducive to undertaking decision-making tasks that necessitate creativity or efficiency, whereas negative emotional states are found to be more adept in contexts that require substantial effort and/or a rigorous analysis of specific issues and their potential ramifications. In a similar vein, they further assert that individuals immersed in a positive mood are prone to adopting excessively optimistic perspectives, there-

by placing undue confidence in their own abilities (Elsbach and Barr, 1999). These findings do not suggest that judges should deliberately strive to adopt a negative emotional state prior to assuming the bench. Rather, they possess the capacity to transcend the inherent tendency to rely on reflective processing when in a favorable emotional state, thereby assuming a more vigilant and circumspect stance. As an illustration, empirical evidence has demonstrated that the mere act of directing them toward mindfulness and holding them accountable for their decisions has the potential to facilitate deeper and more deliberate cognitive processing (Casey, Burke and Leben, 2013).

An additional, equally captivating perspective, which transcends the analysis of the scope of the present study, pertains to the influence of emotions in the formation of the judicial decision-making process. This phenomenon, which remains inadequately explored to date, has often been relegated, possibly due to the deeply entrenched belief that emotions have no legitimate role within the courtroom, where the impartiality of the judge must be upheld as the foundational principle, or owing to the assumption that emotional experiences might compromise this impartiality (Bandes and Blumenthal, 2012). Nonetheless, empirical investigations have shown that judges, contrary to the notion of being detached from emotional influences, are, in fact, subject to the experiential realm of emotions. Furthermore, their judicial duties extend beyond the mere regulation of their personal affective states, necessitating the consideration of the emotions of other participants within the judicial sphere. In this context, Edwards and Miller contend that judges are frequently subjected to negative emotional stimuli, such as anger, sadness, and fear, which arise as reflections of the emotions conveyed by others within the courtroom. This emotional transference, under certain circumstances, may lead to secondary traumatic stress, particularly when judges are consistently confronted with distressing narratives, thereby exerting a direct influence on the cognitive and evaluative processes integral to decision-making (Edwards and Miller, 2019).

The experience of undesired emotions and the subsequent effort to suppress them impairs memory, and for judges, this process may deplete the cognitive resources necessary for effective decision-making (Snider, Devereux, and Miller, 2021). In this regard, the reference to the *theory of dual processing* becomes particularly salient, wherein the system is delineated into the reflexive system and the reflective control system. According to Lieberman, the former of these subsystems is predicated upon primal patterns that arise from the dynamic and continuous interaction between the individual and their environment. These patterns, refined through the inexorable passage of time, evolve into entrenched cognitive frameworks that enable the brain to process information in an automatic, rapid, and efficient manner. Conversely, the SRC is grounded in deliberative engagement and conscious exertion directed toward the execution of tasks (Lieberman, 2003). The reflexive system possesses (RS) the capacity to process information continuously and intermittently, predominantly operating through components of associative memory. In contrast, the reflective control system (RCS) exhibits a more constrained capacity, focusing primarily on working memory and the engagement of analytical thought (Serra, 2021). Consequently, the brain exhibits a notable resistance to its activation. This principle of least effort asserts that decision-makers predominantly rely on the automatic retrieval of pre-existing cognitive schemas to process incoming information, engaging the RCS solely

when compelled to do so—such as when faced with the challenge of acquiring a new skill or resolving a complex issue (Evans, 2006). The issue at hand arises from the fact that, at times, the foundational schemas are constructed upon erroneous or only partially accurate information or are applied inappropriately. Clear instances of schemas that may lead to inaccurate decisions include heuristics and implicit biases, both of which possess the capacity to distort perception, thereby influencing judgment.

Heuristics are conceptual frameworks grounded in a specific subset of available information, thereby facilitating the expedient adoption of decisions (Dale, 2015). These cognitive mechanisms are employed whenever judgments are informed by intuitive recognition, or when the deliberate constriction of the information processed is undertaken. In particular contexts, reliance upon such processes can yield judgments and resolutions of greater precision than those derived from more rational, methodical models. Nevertheless, heuristics are not immune to fallibility; they may lead adjudicators to hastily arrive at conclusions and to make errors in navigating the intricate dilemmas inherent in the deliberative process. As a particularly salient illustration, in a series of empirical investigations conducted with German judges, Englich, Mussweiler, and Strack sought to determine whether decisions pertaining to the imposition of criminal sentences could be shaped by anchors that the participants were aware in advance held no substantive relevance. These anchors were introduced through three distinct modalities: a) a question posed by a journalist; b) a sentencing request from the prosecutor that was explicitly arbitrary; and c) a sentence demand issued by the prosecution, following a roll of dice manipulated by the judge themselves. In a consistent pattern across all scenarios, the decisions were found to be influenced by these seemingly irrelevant anchors (Englich, Mussweiler and Strack, 2006). In a subsequent investigation, the same researchers elucidated that participants exposed to a high anchor reacted more expeditiously to incriminating evidence than to exculpatory evidence. This finding, which implies a primacy of internal coherence in the decision-making process, is further complemented by a more profound insight: despite their expertise, judges exhibited a heightened certainty regarding their decisions, thereby suggesting a disconcerting possibility: experts, in assuming a heightened level of knowledge, may be prone to the fallacy of considering themselves less vulnerable to the biased influences that disrupt the trajectory of their judgments, as though their proficiency insulates them from the subtle forces that govern human cognition (Casey, Burke and Leben, 2013). Casey, Burke, and Leben emphasize the reliance on small, non-representative samples as a heuristic that warrants further scrutiny. In this context, there exists a notable human propensity to regard small samples as reliable reflections of the entire population, a cognitive fallacy that prompts individuals to adjust their expectations based on these circumscribed patterns (Casey, Burke, and Leben, 2013). To elucidate this concept, consider the scenario of a judge who is tasked with rendering numerous decisions throughout the day. Within the depths of their cognitive processes lies an inherent tendency to underestimate the occurrence of continuous sequences within ostensibly random contexts. Guided by this predisposition, and paradoxically, a judge may demonstrate an aversion to treating subsets of individuals—such as four, five, or even six consecutive cases—as overly homogeneous, even though such subsets are constructed retrospectively from the events at hand (Gilovich, Vallone and Tversky, 1985). This hypothetical representation reinforces the classical conception that heu-

ristics sacrifice accuracy in the interest of efficiency, which in turn, may lead to the expansion of implicit biases (Thirsk *et al*., 2022). This construct is characterized as a manifestation of errors or distortions grounded in stereotypes or attitudes that operate at a subconscious level. An attitude is defined as an inherent association between a specific concept and an evaluation, whether favorable or unfavorable. In contrast, a stereotype is intrinsically linked to the association of a given concept with a particular characteristic or attribute. While these two constructs are undoubtedly intertwined, it is of paramount importance to delineate a distinction between them, as a favorable attitude does not preclude the existence of negative stereotypes, and conversely, the opposite scenario remains equally plausible. To exemplify this phenomenon, one might conceive of an individual who holds a positive attitude toward African Americans, yet continues to subconsciously associate this group with weapons. Similarly, it is conceivable that an individual might nurture a favorable stereotype toward Asian Americans, regarding them as particularly adept in the field of mathematics, while concurrently maintaining a globally negative disposition toward their broader identity (Kang *et al*., 2012).

Researchers utilize a diverse array of methodologies to assess implicit bias, with the Implicit Association Test (IAT) being the most widely utilized. This test, a classification task, quantifies the temporal discrepancies between pairings that align with conceptual schemas, represented by words or images. In an effort to investigate the presence of implicit racial biases within judicial contexts, Rachlinski and Johnson administered this test to judges serving in lower courts. Their findings revealed a pronounced bias favoring white individuals in the outcomes obtained from judges within this cohort. In contrast, African American judges did not exhibit a distinctly discernible bias, thus presenting, in the aggregate, a profile that is more nebulous and insubstantial (Rachlinski *et al*., 2009). Further empirical evidence has indicated that offenders possessing Afrocentric physical traits, when convicted of crimes against white victims, were subjected to more severe penalties than those who committed similar offenses against victims of the same racial background (Bagaric *et al*., 2021).

The automatic and influential nature of implicit biases is unveiled with an unsettling degree of clarity, owing to advancements in neurotechnology. Recent studies have employed high-precision neurotechnological instruments to explore the activation of pivotal cerebral regions, including the amygdala, the anterior insula, and the anterior temporal lobe, each of which plays an indispensable role in the processing of perceived threats, instinctive responses of disgust, and the consolidation of social stereotypes, respectively. These structures demonstrate heightened reactivity, which is found to correlate with the extent of unconscious bias associated with racial matters (Salmanowitz, 2017). This finding constitutes merely one of the myriad facets in which these advancements serve as a guiding beacon amidst the shadows that enshroud human comprehension. From this apex, previously uncharted avenues unfold, directing towards a deeper understanding of the underlying mechanisms that shape implicit biases, while concurrently, strategies of meticulous design for their mitigation in the realm of criminal sentencing increasingly come into sharper focus.

3. Neurotechnological pathways to mitigating implicit racial bias in criminal sentencing

Empirical research provides robust evidence to support the notion that implicit biases may play a decisive role in the decision-making process culminating in the imposition of a guilty verdict upon a defendant. This phenomenon, concealed within the recesses of consciousness, appears to weave invisible threads that insidiously erode the objectivity of judicial decisions, influencing biases that transcend the explicit awareness of the judge, yet leaving an indelible imprint upon the final judgment. Within the extensive realm of strategies devised to mitigate this insidious and detrimental influence, the proposal advanced by Salmanowitz emerges prominently, encompassing the following interventions: a) enhancing the awareness of judges; b) subjecting them to a stringent assessment via the Implicit Association Test (IAT) prior to trial; c) attenuating stereotypical associations; and d) the potential incorporation of computerized cognitive training tasks alongside non-invasive brain stimulation techniques (Salmanowitz, 2015; Salmanowitz, 2016). The propositions concerning the enhancement of awareness are firmly grounded in the relationship between motivation and behavior, positing that, should courts succeed in fostering a more acute recognition of the implicit biases residing within judges, they would be compelled—almost as an inherent duty—to recalibrate and temper their actions. Negowetti concurs with this perspective, emphasizing that, although the reliance on entrenched schemas presents a formidable challenge to overcome, implicit biases originating from social categories and schemas are not immutable; on the contrary, they are subject to eradication, provided that one acknowledges the potential influence of conceptions related to race, gender, sexual orientation, and other social categories upon the decision-making process. This dynamic necessitates a departure from the irrational paradigm that perpetuates adherence to predetermined schemas, thereby fostering, instead, a more rigorous and discerning examination of the particular circumstances at hand (Negowetti, 2015). In other words, the authentic safeguard against the peril of issuing subjective, personal judgments does not reside in the denial of the constraints imposed by one's own perspective, but rather in the comprehensive acknowledgment and acceptance of such constraints (Minow and Spelman, 1988). It is of paramount importance to underscore that this internalization is intricately interwoven with the elucidation of the underlying internal mechanisms that govern the decision-making process; indeed, it is within this very realm that the pivotal role of neurotechnological instruments assumes prominence, with their potential unfurling as a transformative tool capable of unlocking novel avenues in the comprehension of the complex cognitive processes that shape our choices.

In accordance with this framework, an exhaustive examination of studies based on fMRI has established a link between amygdala activity and the manifestation of racial biases, thereby positioning it as a neurobiological epicenter for such interactions. This analysis transcends the mere identification of the aforementioned correlation, venturing into the hypothesis that the processes occurring within this cerebral region may be a consequence of the perception of a threat, the roots of which lie in deeply entrenched, profoundly negative cultural associations with Black men and other historically marginalized groups (Chekroud *et al*., 2014). One of the most consequential implications of this revelation pertains to the significance of neuronal

activity in relation to behaviors of group inclusion and exclusion. Within this conceptual framework, intrinsic factors of human identity emerge as pivotal elements that not only shape the manner in which individuals perceive but also influence how they assess others based on their affiliation with social spheres to which they belong. When such identity is aligned with a group historically or culturally associated with pejorative attributes, these perceptions and biases are processed within the amygdala, a cerebral region that concurrently houses and molds the contours of fear (Spain and Bradley, 2018). In alignment with this perspective, empirical evidence suggests that an increased frequency of interracial interactions during childhood correlates with a modulation of amygdala responses toward individuals of African descent, whose familiarity is readily identifiable, in contrast to those perceived as unfamiliar. This finding implies that interethnic contact experiences hold the potential to influence neural processes associated with the perception of the "other" (Cloutier, Li and Correll, 2014). The capacity for mentalization, also referred to as theory of mind, endows humans with the ability to infer the emotions, intentions, goals, and motivations of others, thus facilitating navigation through intricate interpersonal interactions. One of the most salient advantages of neurotechnological tools is their capacity to measure this process in real-time. Within this framework, numerous studies have demonstrated that regions such as the dorsomedial prefrontal cortex and the temporoparietal junction undergo activation when individuals deduce the mental states of others, particularly when these interpretations pertain to members of the same social group (Saxe and Powell, 2006). Recent advancements indicate that individuals with frequent interracial contact (e.g., quality interactions with Black individuals for White participants) appear to engage in a mentalization process similar to observing faces perceived as either Black or White (Tzipporah *et al*., 2022). In a comparable manner, mentalization processes possess the inherent capacity to trigger a perceptual threshold that aids in the discernment of social injustices, especially within the context of violent interactions, wherein their heightened function enables not only the identification of such concerns but also their subsequent unraveling with enhanced clarity (Kubota, 2024).

An additional finding resides in the fact that implicit bias is contingent upon self-regulation. The capacity to adjust and govern our actions is an invaluable ability, one that allows for the adept navigation toward the fulfillment of our most lofty objectives. In light of this principle, empirical evidence has substantiated that implicit racial bias induces heightened activity within the anterior cingulate cortex (ACC) when observing individuals perceived as Black, as opposed to those perceived as White, particularly when the faces displayed deviate from what is conventionally regarded as prototypical, that is, when they subvert the normative structures of pre-established racial stereotypes. The aforementioned suggests that cognitive conflict arises from the discrepancies between individuals' implicit biases and their motivations to be, and/or appear, egalitarian. Sensitivity in the ACC and other regions of the brain associated with executive control becomes more pronounced when individuals become aware that the focus of study pertains to race, and when they perceive that their responses may reflect latent racial bias. This phenomenon is further magnified by the intricate interaction of internal factors, revealing that an enhanced motivation to respond without prejudice may, paradoxically, exacerbate cognitive conflict, even in the absence of explicit directives to control such biases. In this regard, when individuals –

particularly in this context, judges – become fully cognizant of being evaluated, the possibility arises for the activation of an intensified self-regulation mechanism, conceptualized as a conscious and deliberate strategy aimed at counteracting the underlying forces that may compromise the integrity of their decision-making process. This discovery, which arises from the utilization of neurotechnological applications, aligns with the second strategy articulated by Salmanowitz: the rigorous examination through the Implicit Association Test (IAT) prior to the trial (Salmanowitz, 2016). According to Larson, not only would such an assessment serve to augment the motivation of judicial decision-makers to attenuate implicit biases, but it would also facilitate the identification of individuals who possess these biases to a considerable degree of intensity, thereby exposing the subtle intricacies of subconscious predispositions that may exert influence over their reasoning (Larson, 2016).

The third strategy advanced by Salmanowitz is directed toward the attenuation of stereotypical associations. As an illustrative example, by reducing the association between African descent men and violent crime, the stereotype in question could relinquish its deep-rootedness and automaticity, thereby shedding the rigidity that typically defines it. The proposed methods to counteract this phenomenon encompass exposure to counterstereotypical exemplars, such as the presentation of portraits of notable historical figures in spaces of prominence, or the inclusion of Black individuals in the administration of the IAT test (Salmanowitz, 2016). The framework surrounding personalization may similarly offer a profound and enriching perspective. With regard to its manifestation in judicial normative interpretations, brain-reading technologies emerge as a significant epistemological avenue that expands the frontiers of legal knowledge, facilitating the decoding of the cerebral architecture underlying mental states. Far from constituting a mere theoretical conjecture, this premise is intricately intertwined with an axiom of criminal law: the normative principles that govern it are grounded in postulates regarding human conduct, reflecting foundational conceptions of choice, volition, free will, and rationality (Nóbrega, 2024). In this manner, rather than rendering judgments based on pre-established group characteristics, a judge, endowed with the extensive insights provided by neurological data, would be positioned to substantiate their verdicts on attributes inherently specific to the individual in question, thereby transcending the constraints of generalizations that limit the potential for a fair and truly objective judgment. Consequently, the enhancement of personalization afforded by such an extensive compilation of data appears to exert a pivotal influence on the manner in which the criminal justice system administers sentences, particularly in relation to the nuanced considerations of mitigation (Denno, 2019). To illustrate this, various deep learning methodologies have been deployed to achieve a more precise diagnosis of schizophrenia. In an initial phase, Plis and colleagues implemented a model that utilized a deep Bayesian network applied to a dataset derived from a substantial cohort comprising 198 patients and 191 healthy individuals. The accuracy attained through 10-fold cross-validation reached an impressive 91% (Plis *et al*., 2014). Subsequently, Pinaya et al. employed a deep belief network (DBN) to analyze features extracted from brain morphometry data, with the objective of distinguishing between healthy controls (N=83) and diagnosed subjects (N=143). The DBN achieved an accuracy rate of 73.6% within the study cohort (Pinaya et al., 2016). Additionally, abnormalities in the resting state have been

documented, further demonstrating the potential of this approach as a biomarker for clinical diagnosis (Olugbemi *et al*., 2023).

An elevated level of diagnostic precision concerning mental disorders holds profound implications for the evaluation of an individual's criminal responsibility, particularly when the individual, afflicted by such a condition, engages in criminal conduct or, conversely, refrains from action, all stemming from an impairment in their cognitive or volitional faculties. In light of this premise, it becomes imperative to undertake a meticulous examination of the neurobiological underpinnings that give rise to the subsequent manifestation of illicit behavior in the aforementioned context. Schizophrenia inflicts profound and lasting effects on the daily functioning of those who suffer from it. These disturbances, when interacting, precipitate a considerable decline in both social and occupational domains, thereby eroding the individual's autonomy. A diagnosis of schizophrenia necessitates, at a minimum, the presence of two distinguishing symptoms, with at least one of these symptoms manifesting as delusions, hallucinations, or disorganized speech. Moreover, behaviors of extreme disarray or catatonia may manifest, alongside negative symptoms, which are evidenced by a markedly diminished emotional expression or pronounced apathy (Luvsannyam, 2022). These behavioral manifestations are construed as the phenomenological expressions of limitations observed in both subjective and objective measures of executive functions, particularly in inhibition and working memory. In this delicate configuration of alterations, it is postulated that individuals affected by such conditions demonstrate a progressively alarming tendency toward the emergence of aggressive acts directed at others. This assertion is substantiated by robust evidence, which affirms an elevated risk of engaging in both violent and non-violent behaviors when compared to neurotypical individuals (Hodgins, Piatosa and Schiffer, 2013). While it is beyond dispute that the identification of these irregularities in the current state does not, in a categorical sense, substantiate their presence at the precise moment of the commission of the offense, objective validation through the deployment of neurotechnologies could serve to enhance the probability that such irregularities manifested at that exact point in time. At this critical juncture, it is pertinent to underscore that the detection of distinctive patterns may suggest the possibility of a continuing trajectory, thereby suggesting a correlation that would not only reinforce but also elevate the corroboration of neurobiological evidence, thus fostering, through individualized assessment, an unprecedented level of objectivity in the decision-making process, one that would be buttressed by unparalleled scientific validation.

Finally, the final strategy posited by Salmanowitz to mitigate implicit racial biases is grounded in the promising application of two distinct forms of neurointervention: computerized brain training tasks and non-invasive brain stimulation techniques. The former, owing to their unique capabilities, hold the potential to preserve the interactive and dynamic essence inherent in mental imagery exercises, whilst simultaneously establishing a standardized visual environment, without in any way diluting the cognitive richness of the experience (Salmanowitz, 2015). In light of these considerations and transposing this conceptual framework to the matter at hand, judges would find themselves immersed within an interactive simulation, inevitably compelled to navigate a scenario that has been meticulously constructed. Such virtual environments, meticulously designed with algorithmic precision, possess the capacity to construct narratives that not only envelop the participant in a sensorially

immersive experience but also function as transformative catalysts for profound cognitive shifts. These shifts are capable of dismantling the implicit preconceptions that inherently structure and condition one's perception of reality. Within this framework, Salmanowitz delineates a scenario in which the adjudicator, prior to commencing the deliberative process, would engage with a virtual reality device for a span of five minutes, immersing themselves in a digital representation wherein they would adopt the identity of an avatar from a racial group distinct from their own. This immersion would transcend a mere technological experiment, evolving into a strategic instrument designed to promote emotional self-regulation, enhance perspective-taking, and ultimately deconstruct the ingrained stereotypes that may unduly influence decision-making within the judicial context (Salmanowitz, 2018). The connection to Negowetti's framework regarding the adoption of perspective is notably compelling. As previously articulated, this paradigm is understood as the capacity to position oneself within the experiential reality of an individual belonging to a social group ethnically distinct from one's own. Empathy, in this context, impels judges to engage in a comprehensive awareness of their own responses to the parties and their respective arguments, thereby enabling them – as well as those tasked with evaluating their decisions – to scrutinize their own underlying assumptions, identify potential blind spots and biases, and actively seek supplementary information when the circumstances necessitate such actions (Bandes, 2011). At a superficial level, this reference may appear to be in direct opposition to the traditional notion of objectivity, largely due to the positivist conception that objectivity is construed as an abstract, disembodied state, devoid of emotional influence. However, rather than subverting this principle, empathy reveals itself as an essential and inescapable element, as genuine objectivity requires the conscious regulation of emotional responses, situated within the broader emotional-cognitive matrix that underpins judicial decision-making (Blix and Wettergren, 2019).

Ultimately, non-invasive brain stimulation techniques, such as transcranial magnetic stimulation (TMS) or transcranial electrical stimulation (tES), with transcranial direct current stimulation (tDCS) being the most widely implemented form, could offer a more direct approach for enhancing self-regulatory capacities and mitigating stereotypical associations. This can be achieved through the stimulation of pivotal neural regions, such as the medial prefrontal cortex, the anterior temporal lobe, and the dorsolateral prefrontal cortex, or through intervention in extensive neural networks implicated in these processes (Lopes, 2021). In a study conducted by Sellaro et al., participants who underwent tDCS demonstrated an enhancement in cognitive control over stereotype activation, resulting in a reduction of implicit negative bias toward an outgroup (Sellaro et al., 2017). Thus, by enhancing the decision-making faculties integral to the exercise of judicial functions, these interventions have the potential to play a critical role in mitigating racial biases, thereby facilitating the realization of the idealized and rational model of legal reasoning. Nevertheless, valid concerns regarding their safety persist, particularly due to the dearth of research clarifying their long-term effects on individuals deemed healthy, thereby necessitating a careful reflection on the enduring consequences of these technologies upon the human cognitive structure (Lopes, 2021). Such profound considerations open an expansive field of inquiries concerning their ethical and legal permissibility, underscoring the imperative for a rigorous examination that encompasses not only the advancements in neuroscience but also acknowledges the philosophical and moral

dimensions intrinsically linked to their application in such sensitive domains as the judicial system.

4. Final thoughts: defining a way forward

The imprint of implicit racial biases unveils a profoundly intricate issue that not only necessitates a rigorous theoretical analysis but also imposes, with an inescapable urgency, the imperative for a comprehensive reevaluation of judicial praxis in the imposition of criminal sentences. The meticulous incorporation of neuroscientific findings, which bring to light the subtle incursion of latent prejudices, exposes a truth that surpasses the classical tenets of impartiality, revealing structural fractures within a legal framework that, despite its aspirational pursuit of justice, remains inexorably shaped by the subterranean currents of subjectivity. This emerging paradigm, which does not relegate neurotechnology to a secondary role, but rather positions it as a transformative catalyst with profound and far-reaching implications, suggests the potential for a more equitable administration of justice. Within this framework, the implementation of strategies such as computerized brain training tasks and non-invasive brain stimulation techniques inaugurates an uncharted frontier, wherein the certainties of criminal law are inexorably interrogated by the revelation of underlying neurobiological processes that, without the judges' full awareness, exert a decisive influence on their decision-making. Notwithstanding this, the emergence of these tools, with their potential for profound structural reformation, is not devoid of complexities, as it unveils a vast array of ethical, philosophical, and legal concerns that necessitate meticulous scrutiny, measured deliberation, and an application that does not obscure the fundamental principles of justice, but rather elevates them in their most pristine form. This approach necessitates continuous scrutiny of its long-term implications, given that its implementation could irreversibly reshape the foundational principles that have, to date, underpinned the exercise of the judiciary. In this regard, the challenge manifests as a delicate equilibrium: the integration of these advancements must occur without compromising the essential ethical and legal prudence, ensuring that, rather than relegating justice to an unattainable ideal, it is transmuted into a tangible reality that not only addresses the evolving demands of society, but also remains steadfastly aligned with the immutable essence of human dignity.

References

Bagaric, M., et al. (2021). The solution to the pervasive bias and discrimination in the Criminal Justice: Transparent artificial intelligence. *American Criminal Law Review*, 59(1), 1-57. https://ssrn.com/abstract=3795911.

Bandes, S. (2011). Moral imagination in judging. *Washburn Law Journal*, 51(1), 2011-2032. https://ssrn.com/abstract=1948258.

Bandes, S., Blumenthal, J. (2012). Emotion and the Law. *Annual Review of Law and Social Science*, 8(1), 161-181. http:// doi.org/10.1146/annurev-lawsocsci-102811- 173825.

Capurso, T. (1998). How judges judge: theories on Judicial decision making. *University of Baltimore Law Forum*, 29(1), 1-13. http://scholarworks.law.ubalt.edu/lf/vol29/iss1/2.

Casey, P., Burke, K., Leben, S. (2013). Minding the Court: Enhancing the decision-making process. Court review: *The Journal of the American Judges Association*, 49, 76-98. https://iacajournal.org/articles/10.18352/ijca.8.

Chekroud, A., et al. (2014). A review of neuroimaging studies of race-related prejudice: does amygdala response reflect threat? *Frontiers in Human Neuroscience*, 8(179), 1-11. http:// doi.org/10.3389/fnhum.2014.00179.

Cloutier, J., Li, T., Correll, J. (2014). The impact of childhood experience on amygdala response to perceptually familiar black and white faces. *Journal of Cognitive Neuroscience*, 26(9), 1992-2004. http:// doi.org/ 10.1162/jocn_a_00605.

Cohen, H., Bodansky, D. (2024). Close relations: International Legal Realism and Cognitive-Behavioral Studies. *Fordham Law Legal Studies*, 4815319, 1-20. http://dx.doi.org/10.2139/ssrn.4815319.

Dale, S. (2015). Heuristics and biases: The science of decision-making. *Business Information Review*, 32(2), 93-99. http:// 10.1177/0266382115592536.

DeGrassi, S., *et al*. (2012). Ethical decision-making: group diversity holds the key. *Journal of Leadership, accountability and ethics*, 9(6), 51-65. https://api.semanticscholar.org/CorpusID:16861899.

Denno, D. (2019). Neuroscience and the personalization of Criminal Law. *The University of Chicago Law Review*, 86(2), 359-401. https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/7.

Devine, P., *et al*. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6), 2012, 1267-1278. https://doi.org/10.1016/j.jesp.2012.06.003.

Donald, B., Bakies, E. (2016). A Glimpse inside the Brain's black Box: understanding the role of Neuroscience in Criminal Sentencing. *Fordham Law Review*, 85(2), 481-502. https://ir.lawnet.fordham.edu/flr/vol85/iss2/5.

Du, Y. (2021). Racial bias still exists in Criminal Justice System? A review of recent empirical research. *Touro Law Review*, 37(1), 1-26. https://digitalcommons.tourolaw.edu/lawreview/vol37/iss1/7.

Edwards, C., Miller, M. (2019). An assessment of judges self-reported experiences of secondary traumatic stress. *Juvenile and Family Court Journal*, 70(2), 7-29. http:// doi.org/10.1016/j.brat.2009.05.007.

Elsbach, K., Barr, P. (1999). The Efects of mood on individuals' use of structured decision protocols. *Organization Science*, 10(1), 181-198. https://www.jstor.org/stable/2640311.

Englich, B., Mussweiller, T., Strack, F. (2006). Playing Dice with Criminal Sentences: The influence of irrelevant anchors on experts' judicial decision-making. *Personality and Social Psychology Bulletin*, 32(2), 188-200. http:// doi: 10.1177/0146167205282152.

Evans, J. (2006). The heuristic-analytic theory of reasoning: extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), 378-395. http:// doi: 10.3758/bf03193858.

Everson, G. (1919). Human element in Justice. *Journal of Law and Criminology*, 10(1), 90-99. https://scholarlycommons.law.northwestern.edu/jclc/vol10/iss1/10.

Garrigan, B., Adlam, A., Langdon, P. (2016). *Brain and Cognition*, 108, 88-97. http://doi: 10.1016/j.bandc.2016.07.007.

Geyh, C. (2014). The dimensions of Judicial impartiality. *Florida Law Review*, 65(2), 493-551. http://scholarship.law.ufl.edu/flr/vol65/iss2/4.

Gilovich, T., Vallone, R., Tversky, A. (1985). The hot hand in basketball: on the misperception of random sequences. *Cognitive Psychology*, 17(3), 295-314. http:// 10.1016/0010-0285(85)90010-6.

Greenwald, A., Banaji, M. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4-27. http:// doi: 10.1037/0033-295x.102.1.4.

Griffin, T., Wooldridge, J. (2006). Sex-based disparities in felony dispositions before versus after sentencing reform in Ohio. *Criminology*, 44(4), 893-923. http:// doi.org/10.1111/j.1745-9125.2006.00067.x.

Hodgins, S., Piatosa B., Schiffer, B. (2013). Violence among people with schizophrenia: phenotypes and neurobiology. *Current topics in behavioral neurosciences*, 5, 329-368. http://doi.org/ 10.1007/7854_2013_259.

Holmes, B., Feldmeyer, B., King, T. (2020). Sentencing departures and focal concerns: The joint effect of race and gender on departures in United States district courts, 2014-2016. *Journal of Crime and Justice*, 43(5), 598-622. https://doi.org/10.1080/0735648X.2020.1730933.

Huhn, W. (2003). The stages of Legal reasoning: Formalism, Analogy, and Realism. *Villanova Law Review*, 48, 305-380. https://digitalcommons.law.villanova.edu/vlr/vol48/iss1/5.

Kang, J., et al. (2012). Implicit Bias in the Courtroom. *UCLA Law Review*, 59, 1-63 https://www.uclalawreview.org/pdf/59-5-1.pdf.

Kubota, J. (2024). Uncovering implicit racial bias in the brain: the past, present & future. *Daedalus*, 153(1), 84-105.

Larson, D. (2016). Fair and implicitly impartial jury: An argument for Administering the implicit association test during voir dire. *De Paul Journal for Social Justice*, 3(2), 139-172. https://via.library.depaul.edu/jsj/vol3/iss2/3.

Lieberman, M. (2003). Reflective and reflexive Judgement Processes: A Social Cognitive Neuroscience Aproach. In Forgas, J., Williams, K., Von Hippel, W. (ed.), *Social Judgements: implicit and explicit processes*. Cambridge: Cambridge University Press, 44-67.

Light, M. (2021). The Declining significance of race in criminal sentencing: evidence from US Federal Courts. *Social Forces*, 100(3), 1110-1141. http:// doi:10.1093/sf/soab018.

Lopes, G. (2021). Cognitive enhancement in the judiciary. *Direito, Estado e Sociedade*, 59, 1-12. http://doi.org/ 10.17808/des.59.1785.

Luvsannyam, E., et al. (2022). Neurobiology of Schizophrenia: a comprehensive review. *Cureus*, 14(4), 1-7. http://doi.org/10.7759/cureus.23959.

Mack, K., *et al*. (2021). Judicial impartiality, Bias and Emotion. *Australian Journal of Administrative Law*, 28(2), 66-82. https://ssrn.com/abstract=4021234.

Maroney, T. (2011). Emotional regulation and Judicial Behavior. *California Law Review*, 99, 1485-1556. https://scholarship.law.vanderbilt.edu/faculty-publications/772.

Meyer, J., Jesilow, P. (1996). Research on bias in Judicial Sentencing. *New Mexico Law Review*, 26(1), 1-26. http://digitalrepository.unm.edu/nmlr/vol26/iss1/6.

Minow, M., Spelman, E. (1988). Passion for Justice. Cardoso Law Review, 10(1), 37-76. https://larc.cardozo.yu.edu/clr/vol10/iss1/6.

Nagel, S. (1962). Judicial backgrounds and Criminal cases. *Journal of Criminal Law and Criminology,* 53(3), 333-339. https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=5105&context=jclc.

Negowetti, N. (2015). Judicial decision making, empathy, and the limits of perception. *Akon Law Review*, 47(3), 693-751.

Nóbrega, J. (2024). Transtorno por deficit de atención e hiperactividad (TDAH), Neurociencia y Justicia Penal. Madrid: Dykinson.

Olugbemi, T., et al. (2023). Clinical applications of Deep learning in neurology and its enhancements with future directions. In Abraham, A., Dash, S., Kumar Pani, S., García-Hernández, L. (ed.), *Artificial intelligence for Neurological Disorders*. London: Elsevier, 145-158.

Paterson, D., Pardo, M. (2010). Philosophical foundations of Law and Neuroscience. *Illinois Law Review*, 1211, 1212-1250. https://illinoislawreview.org/wp-content/ilr-content/articles/2010/4/Pardo.pdf.

Pinaya, W., et al. (2016). Using Deep belief network modeling to characterize differences in brain morphometry in schizophrenia. *Scientific reports*, 6, 1-9. http://doi.org/ 10.1038/srep38897.

Plis, W., et al. (2014). Deep learning for neuroimaging, a validation study. *Frontiers in Neuroscience*, 8, 1-11. http://doi.org/10.3389/fnins.2014.00229.

Pyun, J. (2015). When Neurogenetics hurts examining the use of Neuroscience and Genetic evidence in Sentencing decisions through implicit bias. *California Law Review*, 103(4), 1019-1046. Available at: https://lawcat.berkeley.edu/record/1126879?v=pdf.

Rachlinski, J., Johnson, S. (2009). Does unconscious racial bias affect trial judges? *Notre Dame Law Review*, 84(3), 1195-1246. Available at: https://scholarship.law.nd.edu/ndlr/vol84/iss3/4.

Rodríguez, S., Curry, T., Lee, G. (2006). Gender differences in criminal sentencing: do effects vary across violent, property, and drug offenses? *Social Sciences Quarterly*, 87(2), 318-339. http://doi:10.1111/j.1540-6237.2006.00383.x.

Salmanowitz, N. (2018). The impact of virtual reality on implicit racial bias and mock legal decisions. *Journal of Law and the Biosciences*, 5(1), 174-203. http://doi.org/ 10.1093/jlb/lsy005.

Salmanowitz, N. (2016). Unconventional methods for a traditional setting: The use of virtual reality to reduce implicit racial bias. *The University of New Hampshire Law Review*, 15(1), 117-160. http://scholars.unh.edu/unh_lr/vol15/iss1/2.

Salmanowitz, N. (2015). Unconventional methods for a traditional setting: the use of neurointerventions to reduce implicit rational bias in the courtroom. PhD Thesis: Duke University, 1-94. https://www.proquest.com/docview/1718184951?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses.

Saunders, J., Midgette, G. (2023). A test for implicit bias in discretionary criminal justice decisions. *Law and Human Behavior*, 47(1), 217-232. https://doi.org/10.1037/lhb0000520.

Saxe, R., Powell, L. (2006). It's the thought that counts: specific brain regions for one component of Theory of Mind. *Psychological Science*, 17(8), 692-699. https://www.jstor.org/stable/40064437.

Sellaro, R., et al. (2017). Transcranial direct current stimulation. In *Theory-driven approaches to cognitive enhancement*. New York: Springer, 99-114.

Serra, D. (2021). Decision-making: from neuroscience to neuroeconomics – an overview. *Theory and Decision*, 91, 1-80. https://doi.org/10.1007/s11238-021-09830-3.

Snider, K., Devereux, P., Miller, M. (2021). Judge's emotion: an application of the emotion regulation process model. *Psychiatry, Psychology and Law*, 1-18. http:// doi.org/10.1080/13218719.2021.1904454

Spain Bradley, A. (2018). The disruptive Neuroscience of Judicial choice. *UC Irvine Law Review*, 1, 1-29. https://scholar.law.colorado.edu/faculty-articles/1208.

Steffensmeier, D., et al. (1998). The interaction of race, gender, and Age in Criminal sentencing: The punishment cost of being young, black and male. *Criminology*, 36, 763-797. http:// doi.org/10.1111/j.1745-9125.1998.tb01265.x.

Thirsk, L., et al. (2022). Cognitive and implicit biases in nurses' judgement and decision making. *International Journal of Nursing Studies*, 133, 1-14. http://doi.org/10.1016/j.ijnurstu.2022.104284.

Tzipporah, D., et al. (2022). Perceiving social injustice during arrests of black and white civilians by white police officers: an fMRI investigation. *Neuroimage*, 255(3), 1-14. http://doi.org/ 10.1016/j.neuroimage.2022.119153.

Uhrig, N. (2016). Black, Asian and minority ethnic disproportionality in the Criminal Justice System in England and Wales. London: Ministry of Justice Analytical Services.

Veiga, A., Pina-Sánchez, J., Lewis, S. (2022). Racial and ethnic disparities in sentencing: what do we know and where should we go? *The Howard Journal of Crime and Justice*, 62(2), 167-182. https://doi.org/10.1111/hojo.12496.

Yoder, K., Decety, J. (2018). The neuroscience of morality and social decision-making. Psychology, *Crime and Law*, 24, 279-295. https://doi.org/10.1080/1068316X.2017.1414817.

— 13 —

Gender-based algorithmic discrimination and just transition in the framework of the fourth industrial revolution and technological innovation

DENISE ALMEIDA DE ANDRADE
Centro Universitário Christus - Unichristus
E-mail: andradedenise@hotmail.com

MONICA SAPUCAIA MACHADO
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa - IDP
E-mail: msapucaia@yahoo.com

Introduction

The Fourth Industrial Revolution – or Industrial Revolution 4.0 - is so named due to the strength of the changes that have occurred in the wake of the capillarization of the use of the internet, especially from the 1990s onwards, in the most diverse areas of human existence, which justifies talking about a "revolution."

The numerous impacts range from the most basic form of day-to-day communication between friends and family to the reach of advanced technology tools related to space exploration and the treatment of diseases. Such positive or negative repercussions do not affect people equally because the particularities of each person and/or group are maximized in this scenario of the Fourth Industrial Revolution.

At this point, we believe it is essential to defend an accurate look at algorithmic discrimination since a significant part of consumer relations, employment, and financial market trends are associated, if not restricted, in the twenty-first century to the use of algorithms, under promises of various benefits, without, however, a critical and responsible evaluation of their repercussions for vulnerable groups, which are on the margins of creation and development of these tools.

From this reflection, we understand that algorithmic discrimination is already a reality, and that discrimination based on gender, to the detriment of women, has

materialized as one of its biases, which will be the central object of our analyses in this article.

For this reason, it is urgent to think of ways to contain such discrimination and build paths in favor of equality. In this sense, the Just Transition is recommended as a central tool for human and economic development that, in fact, dialogues with reducing inequalities instead of contributing to their deepening.

1. Gender-based algorithmic discrimination and the perpetuation of gender inequalities

Technology is not neutral. Until recently, the twenty-first century has been the stage for what seems like a dystopian future: We hire and fire people based on algorithms' recommendations, define advertising campaigns based on mathematical models, and structure a new business model under the direct influence of artificial intelligence.

The Industrial Revolution 4.0 means the fusion between "the technologies of the physical, digital and biological world" (Schwab, 2016, p. 11). This materializes in a hyperconnected society that is increasingly dependent on the internet and technological tools and has not remained alert and critical to the problems and challenges of this moment. In this sense, we highlight algorithmic discrimination based on gender, which affronts consecrated rights and deepens historical inequalities.

This century represents the beginning of a new social and economic organization involving how we relate and communicate. We understand that the changes brought about by disruptive technologies are undeniable, but we must have parameters and limits. According to Schwab (2016), fighting off technological advances is impossible. However, it is essential to limit how it is developed, protect enshrined rights, and recognize that people are not expendable. It is not a simple task amid discourses that constantly reproduce the untruth that machines are neutral and objective. At the same time, people are driven by emotion and biased and, therefore, less reliable.

O'Neil (2016) says that mathematicians and statisticians have been studying people's desires, movements, and purchasing power for decades, which is done in an attempt to predict credibility and calculate the unique potential of each one, depending on the spaces they occupy in society: students, workers, lawbreakers, etc. Based on this premise, it is important to conceptualize algorithms and highlight their importance. For O'Neil (2020, p.19), the algorithm means:

> [...] nothing more than the abstract representation of some process, be it a baseball game, an oil company's logistics chain, a foreign government's actions, or a movie theater's audience. Whether running inside a computer or in our heads, the model uses what we know to predict responses in various situations. We all carry thousands of models in our heads. They tell us what to expect and guide our decisions. (free translation)

O'Neil (2020) explains that algorithms are always subject to errors because, by their very nature, they are simplifications. For this reason, they cannot embed all the complexity and nuances of the real world in themselves, leaving out, inevitably, some information.

Machines need a human to program them; therefore, algorithmic systems do not escape the cognition of their developers, which leads to the direct manifestation of

their judgments and priorities. Therefore, it is pressing to understand that machines' neutrality is an illusion since models do not emerge as objective entities free of subjectivity; on the contrary, they reveal themselves as entities that mirror the objectives and ideologies of their architects.

> The blind spots of a model reflect the judgment and priority of its creators [...] Here, we see that models, despite their reputation for impartiality, reflect goals and ideologies [...]. Our values and desires influence our choices, from the data we choose to collect to the questions we ask. Models are opinions embedded in mathematics (O'Neil, 2020, p. 22). (free translation).

Thus, even in applying apparently impartial mathematical methods, our own values and desires operate subtly, permeating from the choices of data to be collected to the questions we direct to the model. Therefore, in the end, the creative and programming subjects of the machines will reproduce, even if unconsciously, several of their biases and understanding of the world. Therefore, decisions made by machines, since they were programmed by humans who carry their intrinsic opinions, will not be neutral or bias-free.

Legitimate objectives, noble intentions, laudable projects... These are generally the starting points for a new technology or for the application of a consolidated technology in a non-obvious context.

The denial of biases in the handling of technological tools has always been a maxim among developers and primary users of these instruments. However, this only starts from a comfortable discourse that the punctual action of the "non-human" is neutral and that this is a guarantee of isonomy in favor of meritocracy. It turns out that human beings are partial, fallible, and constituted under the power of prejudices and stereotypes.

Therefore, such human biases can be camouflaged within their codes, that is, within the developed models constituted by sets of assumptions, and some of them can be considered harmful (O'Neil, 2020). The question arises: Is it possible to eliminate these biases even knowing this?

Human preconceived ideas, intrinsic to cognitive nature, can lead to experiences of discrimination, so biases in algorithms would be no different. This fact is based on the premise that these systems, although impersonal in their formulation, are conceived and programmed by individuals susceptible to their own unconscious predispositions.

In the case of system-automated decisions, all are at risk of discrimination. These risks can become a mass phenomenon and quickly lead to cumulative disadvantages over time (Orwat, 2019, p. 22). Therefore, the call to combat discrimination focuses on the rejection of differentiation practices that exceed the limits of reasonableness, and the constitutional and legal frameworks previously established legitimately and democratically. These practices hurt the principle of equality and, therefore, demand an active response to promote equity and the protection of individual rights.

Thus, the choices made during their development can profoundly impact on individual rights and freedoms. An algorithm can be understood as an instruction manual or a detailed recipe that aims to guide the steps to achieve a specific result – for example, solving problems and assisting in decision-making. It is composed of transparent and sequentially ordered instructions, ensuring that, when followed, the desired objective is achieved efficiently and correctly. For O'Neil (2020, p.19), an algorithm:

> [...] is nothing more than the abstract representation of some process, whether a baseball game, an oil company's logistics chain, a foreign government's actions, or a movie theater's audience. Whether running inside a computer or in our head, the model picks up what we know and uses it to predict responses in varied situations. We all carry thousands of models in our Heads. They tell us what to expect and guide our decisions.

We know that algorithms use pre-encoded data to teach them to identify patterns and replicate them in different contexts or tasks. Algorithmic discrimination begins when the input dataset is influenced by biases that introduce distortions to the results, generating outputs that do not correctly reflect the programmer's intentions. Thus, considering that human biases can lead to experiences of exclusion, algorithm biases would be no different.

In the face of the exponential advance of neurotechnology driven by disruptive technologies and generative artificial intelligence, it is vital to recognize that "in the virtual world (apparently neutral) human rights, especially those who are part of vulnerable groups, can be even more subject to potential damage, given the presence of prejudices and biases in the digital environment, perpetuating historical inequalities" (Andrade *et al*., 2024, p. 239) (free translation).

On the same note:

> The fact that few enjoy the benefits of this technological shift few enjoys should not be surprising. On the contrary, it is the result of marginalization processes that, combined, intensify structures and practices that should not be perpetuated and that need to be faced. This confirms the importance of efforts such as that seen in the elaboration of the Inter-American Legal Declaration on Neuroscience, Neurotechnologies, and Human Rights (Andrade and Bertolin, 2022, p. 86) (free translation).

Algorithmic discrimination does not occur only based on gender but also the basis of race, socioeconomic condition, etc. We highlight a case of a Russian subsidiary of an American company. Xsolla specializes in payment software solutions for the online gaming industry, which fired, in August of 2021, 150 employees out of a total of 450 on the recommendation of an algorithm. The math model named the employees "little committed" and "unproductive" under the justification that the company would have stopped growing by 40%. We noted, however, that none of the 150 employees, according to the Russian Forbes, were in a senior position.

This episode is not unique. In 2019, the giant Amazon acted likewise when it fired several employees based on computer criteria without the reflection or contribution of the director of operations, the human resources sector, or any other team responsible for staff matters (Echarri, 2021).

Both occurrences are not classic cases of discrimination based on gender. However, we know that formal work opportunities are not equal for men and women for several reasons, among which we highlight the workload that women dedicate to care activities and motherhood. So, even in cases where there is no explicit discrimination against women, it is necessary to analyze the repercussions since they affect them more strongly.

At the same time, gender is one of the most relevant variables in explaining the setbacks in the "digital world". That said, two concepts deserve to be highlighted: gender equality according to UNESCO and discrimination against women by the United Nations (1989):

> Gender Equality is one of UNESCO's two global priorities. In this understanding, gender equality refers to the equal rights, responsibilities, and opportunities of women and men and girls and boys. It implies that

> the interests, needs, and priorities of both women and men are taken into consideration, recognizing the diversity of different groups of women and men. Gender equality is a human rights principle, a precondition for sustainable, people-centered development, and a goal in and of itself (UNESCO).

According to the Convention on the Elimination of All Forms of Discrimination against Women:

> Discrimination against women shall mean any distinction, exclusion, or restriction made on the basis of sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment, or exercise by women, irrespective of their marital status, on the basis of equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field. (UN, 1979).

It is important, then, to present cases that unequivocally demonstrate gender-based algorithmic discrimination.

We have reports of linking negative biases on platforms related to Black, Latina, and Asian women, correlating them with pornography and reprehensible social behaviors. In 2010, a case of algorithmic discrimination was reported when Nikon cameras were not able to recognize Asian faces, sending the indication that the user was blinking (Matamoros-Fernández and Farkas, 2021).

LinkedIn, the world's largest professional network that connects people with opportunities and peers, discovered that hiring algorithms referred more men than women to available positions due to the "more aggressive" profile in the job search on the platform (Frasão, A., and Mulholland; C., 2021). Cases of hiring and firing based only on algorithm recommendations are increasingly common nowadays.

LinkedIn's case directly affects women's employability and their insertion or reinsertion in the formal labor market. In the situations of Xsolla and Amazon, the dismissals were carried out without considering the peculiarities that women worldwide have taken on most of the care of children, the elderly, the sick, and the home. In the case of Xsolla, we have to remember that it took place during the COVID-19 pandemic when children went homeschooling, and women were the ones who took care of them.

Another example of algorithmic bias involves Google, which led to acts of discrimination in 2009. When typing "black girls" on Google's search engine, the results were associated with pornographic content. Episodes like this reproduce the marginalization of women and other minorities in society, a fact that caught the attention of Safiya Noble, who ended up culminating in the term "algorithms of oppression" (Noble, 2021).

Gender-based algorithmic discrimination is particularly concerning because the myth of technological neutrality fuels it. This myth deepens inequalities and disregards particularities and exceptionalities inherent to women's human condition under the premise of falsehood.

> The fact that few enjoy the benefits of this technological shift should not be surprising. On the contrary, it is the result of processes of marginalization that, combined, intensify structures and practices that should not be perpetuated and that need to be faced. This confirms the importance of efforts such as that seen in the elaboration of the Inter-American Legal Declaration on Neuroscience, Neurotechnologies, and Human Rights. (Andrade and Bertolin, 2022, p. 86) (free translation)..

Joy Buolamwini and Timnit Gebru (2018) point out that "facial recognition systems show higher error rates for women and dark-skinned people," demonstrating that vulnerabilities, in this case, gender and race, intersect and potentialize harm.

The discriminatory action resulting from the operation of these algorithms is not, as the companies that own the control of the codes allege, the result of the behavior of their users. On the contrary, the results derive from a business model created and structured for this, based on the concealment and naturalization of the prejudices of its programmers, in a context of "cyber utopian" visions.

For O'Neil (2020, p. 29), we have to know and combat the so-called "opaque and invisible algorithms," about which there is little transparency in how they were built and how they work or operate. Even wrong or harmful, their models were beyond challenge, and "they tended to punish the poor and oppressed of society while further enriching the rich." (O'Neil, 2016, p. 7).

In this sense, it is urgent to understand that there is a vast difference between fair and biased algorithms. This understanding is crucial and lies in the intentionality and transparency behind the algorithm's design and implementation. Fair algorithms are crafted with an awareness of potential biases and include mechanisms to counteract them. This process often involves scrutinizing the data, adjusting for known biases, and regularly auditing the algorithm's outcomes (O'Neil, 2016). On the other hand, biased algorithms, intentionally or unwittingly, ignore these disparities, leading to results that unfairly disadvantage certain groups.

Although the advances in the affirmation of rights and expansion of opportunities, such as the growth of the labor market for women, for Wajcman (2012), despite being perceived as a driving force, technology, and its gains have not guaranteed the establishment of a new order of organization of States and societies that would be less unequal. This is because it is observed that the technology market is built based on pre-existing relations of sex, class, and race, which shape the structure of the workforce and employment opportunities.

Therefore, it is urgent to discuss and build new paths to protect rights and promote gender equality, such as Just Transition, to be used as a tool to balance forces in such a complex scenario.

2. Just Transition and Gender Equality

Just Transition is a concept that arises from the need to adjust economic and environmental policies to address climate change, ensuring that its impacts and benefits are distributed fairly and equitably among different social groups, economic sectors, and regions. This concept is particularly relevant in the context of a shift from a fossil-based production model to a low-carbon model, which aims to promote more sustainable development.

Henar Álvarez Cuesta (2020) points out that the challenge of combating climate change must be approached not only from an environmental point of view but also from a labor and social perspective. This involves the transformation of economic sectors that directly impact people's lives and jobs, with strategies that provide both mitigation of climate change effects and adaptation to these effects.

The origin of the term "just transition" dates to the 1990s, when US trade unions began developing the concept as a support program for workers who lost their jobs due to environmental protection policies, initially focused on sectors such as coal (Smith, 2017).

However, the concept has evolved into a more comprehensive approach. Trade unions and the like have come to interpret "just transition" as an institutional/state effort to plan and invest in the Transition to jobs, sectors, and environmentally and socially sustainable economies. As the understanding of the climate crisis deepened, "just transition" was explicitly linked to actions against climate change, and initiatives were taken to incorporate this idea into international regimes, including negotiations under the Convention-United Nations Framework on Climate Change (UNFCCC).

This development culminated in the inclusion of the concept in the preamble to the 2015 Paris Agreement, where countries recognize the importance of considering the imperatives of a fair labor force transition and the creation of decent work and quality jobs according to nationally defined development priorities.

Despite the concept's solidification in the international environmental context, the idea of a Just Transition advocated in this article broadens the scope of this understanding. A just Transition must ensure that people, countries, and regions are protected against the deepening of existing social and regional inequalities and that governments and corporations that profit from these changes plan and bear the cost of leaving no one behind.

In addition to the transformations caused by more sustainable and efficient production models, we are facing today the changes in the productive structure brought by the 4.0 revolution, especially by using artificial intelligence (AI) in the productive process. These disruptive phenomena - the change of the energy matrix and revolution 4.0 - must be accompanied by plans for a fair transition so that they do not deepen the existing social and economic inequality abyss and are tools for advancing social justice.

When considering social inequality and the labor market, the gender issue is especially critical. In 2024, women still earned, on average, 77% of what men did for the same work, and the situation is even more unfavorable for women with children (UN Women, 2024). The gender wage gap is a persistent reality in the capitalist system and has been an integral part of the structure of the global labor market. Although there are policies that seek to value women's work equally to men's, the data indicate that advances are slow and often ineffective.

> Since 1971, the number of countries that have adopted pay equity laws has increased from 2 to 98, with regional differences in uptake of such laws. However, only one in five economies legislating equal pay for equal work has also implemented mechanisms to redress the pay gap. (UN Women, 2024, p. 36).

In this structurally unequal scenario, disparities become even more evident in STEM (science, technology, engineering, and mathematics) fields. Women make up only 28.2% of the STEM workforce, while they represent 47.3% in non-STEM sectors. In leadership positions within STEM fields, critical to the future of the global economy, female representation drops to a mere 10% (WEF, 2024).

In the context of disruptive transformations in the production process, such as the energy transition and the Industrial Revolution 4.0, technology assumes a central role in labor market innovations. The skills and knowledge from the field of STEM are increasingly valued, concentrating on employment and income opportunities. However, women face significant disadvantages in this scenario, as historically, STEM fields have been dominated by men due to socioeconomic, cultural, and educational factors (Chaubey, 2023).

A crucial distinction between the Industrial Revolution 4.0 and the changes required for the energy transition, compared to previous industrial revolutions, lies in the current ability to anticipate the impacts of these changes on society and mitigate some of the disturbances caused by these transformations. In this regard, the Organization for Economic Co-operation and Development - OECD has produced a series of studies on the impact of artificial intelligence on various sectors, including the labor market. The results of these studies often highlight a significant disadvantage for women. In the October 2024 document titled "WHO WILL BE THE WORKERS MOST AFFECTED BY AI?", the OECD concludes that the most affected will be workers without higher education, women, and older workers (OECD, 2024).

It is interesting to note that although women represent the majority of university graduates in several countries, such as the United States (Pew Research Center, 2024), in most European countries (EUROSTAT, 2023) and Brazil (IBGE, 2022), the research indicates that workers without higher education will be more impacted, referring indirectly to men without higher education, since regardless of academic training, women, having or not higher education, are among the most vulnerable groups to the effects of artificial intelligence on the labor market.

In the context of energy and ecological transition, productive remodeling that promises new job opportunities and new fields of action is predominantly based on STEM knowledge. Women whose professional areas are impacted by the necessary energy transition often have skills less aligned to new technologies. Therefore, the emerging scenario indicates that the two disruptive changes in the production model of the twentieth century tend to affect workers in functions principally exercised by women, such as those that are less technological, more focused on support, organization of spaces, work processes, and customer services.

The Just Transition is a necessary concept for promoting structural public policies of social justice, which aim to combat historical inequalities and prevent the deepening of existing inequalities.

In that context, historical inequalities and wage gaps are widening, already disproportionately affecting marginalized groups, with women often at a disadvantage. Artificial intelligence and energy transformation, as disruptive forces, have the potential to both perpetuate and amplify these disparities, requiring an integrated approach to social justice that considers redistribution, recognition, and representation to mitigate these impacts.

Nancy Fraser (2002), in her analysis of social justice in globalization, highlights the importance of approaching these issues through a three-dimensional perspective that emphasizes the redistribution of resources, the recognition of marginalized identities, and the need for equal participation. Fraser (2002) argues that the struggles for recognition often end up displacing the struggles for economic redistribution, which can weaken the fight against material inequalities. True social justice must integrate these dimensions, ensuring that all people can participate as equals in social interactions, which is crucial to counteract economic recognition and equal participation disparities.

Artificial intelligence systems represent a challenge in technology, as they can inadvertently or purposefully perpetuate existing prejudices. As Sloan and Warner (2020) identified, proxy variables used in AI algorithms, such as credit scores to as-

sess safe driving ability, may reinforce discrimination based on class, race, and gender. This practice can have disproportionately negative impacts on women and other minorities, who historically face disadvantages in accessing financial resources and opportunities.

These disruptive changes present challenges to states and organizations that think about public policies and control or monitoring measures; some countries and international organizations have proposed parameters to regulate these two significant changes in the production model. However, when it comes to the energy transition, changes are much more focused on implementing policies to reduce carbon emissions than regulations to ensure a fair transition of these policies.

Despite this, in 2021, the European Union established the Just Transition Fund—JTF—an initiative created to mitigate the socioeconomic impacts associated with transitioning to a sustainable and climate-neutral economy. As part of the Just Transition Mechanism, which integrates the Investment Plan for a Sustainable Europe, JTF focuses on supporting regions and workers disproportionately affected by the changes needed to achieve these environmental goals.

The JTF seeks to facilitate the diversification and modernization of local economies to soften the negative employment repercussions caused by removing industries that are dependent on fossil fuels or are carbon intensive. By retraining workers and fostering entrepreneurship, the fund aims not only to protect existing jobs but also to create new opportunities in sustainable sectors.

It is important to note that the fund recognizes that energy and technology transition may exacerbate pre-existing disparities. The fund's regulation emphasizes that promoting women's access to the labor market, equal pay, and female entrepreneurship is essential to ensuring equity during and after the Transition. This approach ensures that transition policies address climate challenges while promoting a more just and equitable society.

The fund's resources are directed to the most affected regions, supporting projects that promote clean technologies, energy efficiency, and emission reduction. By integrating social justice and climate policy, the JTF exemplifies how energy transition initiatives can and should address both the immediate economic consequences and the long-term development needs of impacted communities.

Regarding Industrial Revolution 4.0, in particular, when it comes to the use of algorithms and artificial intelligence for decision-making, the European Union has adopted a proactive approach to regulation, trying to establish harmonized rules for security, ethics, and human centrality in the development and implementation of AI according to the Regulation (EU) 2024/1689. This initiative focuses on protecting the health and safety of citizens. It emphasizes the importance of promoting fundamental rights and gender equality, ensuring that AI is used as a force for good, and promoting innovation and employment fairly and inclusively.

A critical point in this regulatory framework is the concern with perpetuating gender disparities, especially in areas related to employment. Automation and technological transformations induced by AI pose significant risks of exacerbating existing inequalities, strengthening gender stereotypes, and limiting women's access to emerging opportunities. Thus, regulation seeks to mitigate these risks by ensuring

that AI technologies are developed and implemented in ways that do not reinforce social and gender inequalities.

In September 2024, 57 countries, including the European Union, the United States, and the United Kingdom, signed the first legally binding Artificial Intelligence (AI) Convention. Under the aegis of the Council of Europe, the AI Convention was adopted after extensive discussions that began in 2022, highlighting the growing international concern about AI systems' ethical and social impacts. This treaty imposes on signatory countries the responsibility to monitor the development of AI to ensure that its use respects strict parameters of public protection, private data, human rights, democracy, and the rule of law.

The Convention recognizes AI's potential benefits in promoting human prosperity, sustainable development, and gender equality, including the empowerment of women and girls.

In Latin America, according to the Regulatory Mapping report on Artificial Intelligence in Latin America (2024), the regulation of Artificial Intelligence (AI) is at different stages of development, ranging from legislation approved to bills in progress.

Argentina is at the forefront, with a specific legal framework for AI. The country has developed several strategies and action plans, including the National Plan for Artificial Intelligence, which guides measures to maximize AI development and implementation opportunities.

Chile is developing a national AI policy integrating technology into key sectors such as health, education, energy, and the environment. The Chilean government is promoting AI as a tool for economic and social development, with a critical look at ethics and social impacts.

Colombia is also moving forward with AI legislation, with bills that discuss governance, ethical use, and security measures for AI systems. The country seeks to integrate AI responsibly into its digital ecosystem. Both Mexico and Uruguay have bills under discussion that aim to establish guidelines for the development and use of AI, focused on ensuring that technology is used ethically and safely and protecting citizens' rights.

In Brazil, the regulation of Artificial Intelligence (AI) has advanced. Bill 2.338/2023, which has been passed in the Senate, establishes general national standards for developing, implementing, and responsible use of AI systems. This project emphasizes the protection of fundamental rights and the implementation of safe and reliable systems, respecting the centrality of the human person and respect for human rights, including equality and non-discrimination.

Despite considerable advancements in the regulatory frameworks concerning artificial intelligence (AI) across various regions, including the pivotal steps taken by the European Union, United States, and United Kingdom with the signing of the AI Convention, a fully consolidated legal framework regarding AI use remains elusive, particularly concerning policies to mitigate its adverse effects on people's lives.

The AI Convention marks a stride in international law, aiming to monitor AI development rigorously and ensure its alignment with human rights, democracy, and the

rule of law. However, the convention also highlights critical gaps, notably in addressing the specific needs and vulnerabilities of different groups, including women.

In Latin America, where countries like Brazil have made strides in creating national guidelines for AI, the focus has predominantly been on fostering innovation and managing privacy and data protection without explicitly addressing how these technologies could perpetuate gender disparities or affect employment among vulnerable populations.

Conclusion

Automation, artificial intelligence, and mass data processing can become powerful tools for deepening inequalities, such as excluding or reducing women's chances of entering the formal labor market.

There is an attempt to approximate arbitrary decisions of scientificity under the arguments of neutrality and objectivity, both untruths since everything is produced from a human component and consequently has biases.

The mathematization of complex aspects of human life is already causing the escalation of inequalities and perpetuation of discrimination, such as those that occur based on gender.

Given the transformations caused by the energy transition and the Industrial Revolution 4.0, the need for a just transition becomes even more urgent to avoid worsening social and regional inequalities. In addition, to ensure that no one is left behind, this Transition must be planned and funded by governments and corporations that benefit from these changes.

However, for social justice to be effectively promoted, it is essential to address structural challenges—such as gender inequality in the labor market, especially in STEM areas where women remain underrepresented. Thus, inclusion and equity policies must be prioritized so that technological innovation and economic sustainability can coexist with social justice, ensuring real opportunities for all.

We point to Just Transition as one way to contain the indiscriminate use of technology, especially the uncritical management of algorithms, which has led to algorithmic discrimination. Nevertheless, for Just Transition to be effective, governments, businesses, and social organizations must work together to plan and implement policies that guarantee decent work opportunities and sustainable development, promoting social justice and environmental protection in an integrated way.

References

Alvarez, H. (2020). Transición justa y lucha contra el cambio climático en el Pacto Verde europeo y en el Proyecto de Ley de Cambio Climático en España. Iuslabor. *Revista D'Anàlisi de Dret del Treball*, 2, 74-100, 28 jul. Universitat Pompeu Fabra. DOI: http://dx.doi.org/10.31009/iuslabor.2020.i02.04.

Andrade, D. A. de; Bertolin, P. T. M. (2002). Neurotecnologias, estereótipos de gênero e relações de consumo: uma reflexão necessária. In: Lopes, A. M. D. (ed.) *Neurodireito, neurotecnologia e direitos humanos.* Porto Alegre: Livraria do Advogado, p. 83-94.

Andrade, D. A. de; Baima, A. E. B.; Araújo, C. R. B. (2024). A face invisível da inovação, desafios da neurotecnologia e dos neurodireitos no século XXI: uma análise a partir da discriminação algorítmica. In: In: Lopes, A. M. D. (ed.) *Desafios da interface neurodireito e direito artificial,* p. 233 - 254.

Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In: *Conference on fairness, accountability, and transparency*. PMLR, p. 77-91.

Chaubey, P. (2023). Industrial Revolution 4 and Indian Women in STEM: Challenges and Prospects. Journal of National Development, 36 (1). https://jndmeerut.org/wp-content/uploads/2024/01/2-2.pdf

Echarri, M. (2021). 150 demissões em um segundo: os algoritmos que decidem quem deve ser mandado embora. *El País.* https://brasil.elpais.com/tecnologia/2021-10-10/150-demissoes-em-um-segundo-assim-funcionam-os-algoritmos-que-decidem-quem-deve-ser-mandado-embora.html

Eurostat. (2023). More women than men held tertiary degrees in 2022. *Eurostat*. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20230530-3#:~:text=In%202022%2C%2042%25%20of%20the,compared%20with%202021%20(41%25).

Frasão, A. and Mulholland, C. (Host). (2021, October). Discriminação Algorítmica de gênero (7). [Audio podcast episode]. In *Spotfy*. https://open.spotify.com/episode/5mLXHFn88t6Z64uCNWD1nB?si=bFi6OsSvRDqgqPV16z6Prg

Fraser, N. (2002). A justiça social na globalização: redistribuição, reconhecimento e participação. *Revista crítica de ciências sociais*, (63).

IBGE - Instituto Brasileiro de Geografia e Estatística. (2022). *Estatísticas de Gênero: Indicadores Sociais das Mulheres no Brasil.* https://biblioteca.ibge.gov.br/visualizacao/livros/liv102066_informativo.pdf

Lane, M. (2024). Who will be the workers most affected by AI?: A closer look at the impact of AI on women, low-skilled workers, and other groups. *OECD Artificial Intelligence Papers*, (26), OECD Publishing, Paris. DOI: https://doi.org/10.1787/14dc6f89-en.

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, Hate Speech, and Social Media: A Systematic Review and Critique. Television & New Media, 22(2), 205-224. https://doi.org/10.1177/1527476420982230

Noble, S. U. (2021) *Algoritmos da opressão*: como a Google fomenta e lucra com o racismo. Santo André: Rua do Sabão Ed..

O'Neil, C. (2016). *Weapons of Math Destruction*: how big data increases inequality and threatens democracy. UK: Penguin Random House.

O'Neil, C. (2020). *Algoritmo de Destruição em Massa*. Santo André: Editora Rua do Sabão.

ONU Mulheres. (2024). Women in the changing world of work Facts you should know. *UN Women.* https://interactive.unwomen.org/multimedia/infographic/changingworldofwork/en/index.html

Orwat, Carsten. Risks of Discrimination through the Use of Algorithms A study compiled with a grant from the Federal Anti-Discrimination Agency. *Institute for Technology Assessment and Systems Analysis (ITAS) Karlsruhe Institute of Technology (KIT).* https://www.antidiskriminierungsstelle.de/EN/homepage/_documents/download_diskr_risiken_verwendung_von_algorithmen.pdf?__blob=publicationFile&v=1.

Pew Research Center. (2024). U.S. women have outpaced men in bachelor's degree completion for decades. *Pew Research Center*. https://www.pewresearch.org/short-reads/2024/11/18/us-women-are-outpacing-men-in-college-completion-including-in-every-major-racial-and-ethnic-group/sr_24-11-18_college-gender-gap_1/

Schwab, K. (2016). *A quarta revolução industrial.* São Paulo: Edipro.

Sloan, R. H.; Warner, R. (2020). Beyond bias: Artificial intelligence and social justice. Va. JL & Tech., 24, 1.

Smith, S. (2017). *Just Transition A Report for the OECD*. Bruxelas: Just Transition Centre Ed.

UNESCO. Gender Equality Glossary. *World Heritage Convention*. https://whc.unesco.org/en/glossary/251/

United Nations. (1979). *Convention on the elimination of all forms of Discrimination Against Women*.https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women

Wajcman, J. (2012). Tecnologia de produção: fazendo um trabalho de gênero. *Cadernos Pagu*, Campinas, n. 10, p. 201–256.

World Economic Forum. (2024). *Global Gender Gap 2024*. Genebra: World Economic Forum.

— 14 —

Child influencers and the neuro rights of children and adolescents: an analysis of vulnerability in the digital environment

PATRÍCIA MOURA MONTEIRO CRUZ
Professor at Universidade de Fortaleza
patriciamoura@unifor.br

RENATO VILARDO DE MELLO CRUZ
Prosecutor of Ceará and Lawyer.
renatovmcruz@gmail.com

ANTONIO JORGE PEREIRA JÚNIOR
Full Professor Universidade de Fortaleza
antoniojorge2000@gmail.com

Introduction

The phenomenon of child digital influencers, also known as kidfluencers, has expanded significantly in recent years, transforming into a multi-billion dollar industry that currently moves around US$ 2.9 billion (Abeele *et al*., 2024). This scenario emerges in a context of growing child and adolescent participation in social networks - in Brazil, for example, 88% of children and adolescents between 9 and 17 years old who are internet users have profiles on social platforms, representing approximately 23 million users in this age group (CGI.br, 2024).

The performance of children and adolescents as digital content producers characterizes a new communicational paradigm, in which these individuals simultaneously assume the roles of consumers and producers - the so-called prosumers. This dynamic, while potentially providing opportunities for creative development and socialization, also raises significant concerns regarding the protection of fundamental rights of these individuals, who are in a peculiar condition of development.

The multiple vulnerability that characterizes this phenomenon manifests itself in different dimensions: psychological, due to impacts on socioemotional development; economic, through the potential commercial exploitation of child labor; and neurological, considering the susceptibility to algorithmic manipulation techniques and persuasive design employed by digital platforms. This last dimension gains special relevance given the development of neuroscience and its interface with digital technologies, demanding a new perspective of protection through so-called neurorights.

This study aims to analyze the implications of children and adolescents acting as digital influencers from the perspective of neurorights, investigating how these can be applied to provide more effective protection in the digital environment. Specifically, it seeks to: (i) understand the characteristics and dynamics of the kidfluencers phenomenon; (ii) identify vulnerabilities and risks associated with digital exposure; (iii) analyze the applicability of neurorights in the context of childhood and adolescence; and (iv) propose guidelines for public policies and protection practices.

The relevance of this investigation is justified by the pressing need to establish regulatory frameworks and protection mechanisms appropriate to this new reality, considering that existing legislation does not yet fully address the specificities of the digital environment and its impacts on child and adolescent development. Furthermore, the growing monetization of children's and adolescents' digital presence demands a critical reflection on the ethical and legal limits of this practice.

1. Children as prosumers in the digital environment

The phenomenon of child and youth prosumers on social media represents a significant transformation in contemporary communication and socialization dynamics. This emerging paradigm is characterized by the intersection between digital content consumption and production processes, where young audiences assume a dichotomous position, simultaneously manifesting active and passive behaviors in platform-mediated interactions. This shift reflects a reality where children's expressiveness and representations predominantly respond to a specific cultural model: consumer culture (Tomaz *et al*., 2023).

The concept of prosumer, originally proposed by Toffler (1980), gains new dimensions with the emergence of digital participatory culture. Children, previously considered merely passive content receivers, now actively articulate information consumption and production, taking a leading role in communication platforms (Cruz and Cruz, 2024). This dynamic is particularly evident in the kidfluencer phenomenon.

Kidfluencers emerged as a new category of child celebrities that fundamentally differs from traditional forms of child artistic work. Unlike traditional acting, which requires entertainment industry connections or specific training, social media content production only demands a social media account, good ideas, some filming and editing skills, and luck (Masterson, 2021).

The content produced by these young influencers spans a wide variety of formats and themes. On YouTube, most content from leading child influencers focuses on activities such as electronic games, narratives, product and toy reviews, unboxing, challenges, and tutorials (Martínez Allué and Martín Cárdaba, 2024). These produc-

tions frequently employ standardized expressions at the beginning and end of videos, along with emphatic language and distinctive visual elements that contribute to establishing a unique channel identity.

Participation in digital communities materializes through different forms of interaction, with platforms generally implementing a metrics system including views, reactions, comments, and followers, establishing an engagement dynamic fundamental for profile growth and monetization. This system functions as a form of "fame gamification," where engagement statistics can be accessed through commercial management dashboards, stimulating continuous content production (Albuquerque *et al*., 2024).

Like other social media content, production by young influencers can be monetized through multiple mechanisms. The main revenue streams include financial compensation from companies, such as toy manufacturers paying children to post videos playing with their products, and the Google AdSense program, which enables advertising on YouTube channels with view-based payments (Geider, 2021). Some young influencers achieve significant earnings - cases like Ryan Kaji, who at age ten earned US$30 million in 2020 through YouTube advertising revenue, exemplify this activity's economic potential (Reardon, 2022).

Precisely because of these various monetization possibilities, children and adolescents' participation dynamics on social media frequently transcend artistic and cultural expression limits, entering commercial exploitation territory, making it necessary to distinguish recreational artistic expression from what should be identified as child labor. While the first comprises non-profit activities like dance, theater, and music in school or social settings, the second is characterized by third-party monetization of children's artistic activity, regardless of compensation to the developing artist (Andrade, 2022).

The early professionalization of children's digital presence frequently involves parents' or guardians' active participation as activity managers, sometimes becoming these families' main income source. In some cases, children are intentionally prepared by their parents to become commodities and "human billboards" from birth, in a phenomenon termed "micro-microcelebrities" (Abidin 2020). The economic expression of kidfluencer activities can therefore generate misaligned incentives for healthy parenting and an imbalance between children's rights and wishes and the imposed responsibility of providing family income.

However, one should not look only to families for creating these kidfluencers. Despite significant family participation, specialized agencies like Clock.Work exploit these influencers' activities, developing technically sophisticated advertising campaigns combining native advertising and host selling, often circumventing existing regulations through inadequate or imperceptible disclosures (Feller and Burroughs, 2022).

The economic incentive for content production can result in intense material production volume. The Toys and Little Gaby channel, once the UK's largest kidfluencer channel, posted over 250 YouTube videos in four years, while Ryan Kaji's channel accumulated at least 1,830 videos in five years (Masterson, 2021), quantities raising questions about impacts on involved children's well-being and development.

Content produced and offered to children - whether in domestic, school, or media environments - influences their interactions and, through synthesis, sustains processes that may question or reproduce social realities, enabling both creative ways of reading and narrating the world and generating social constraints. Therefore, risks from exposure to such social media content production and consumption deserve deeper exploration.

2. Vulnerabilities in the digital environment

Children and adolescents' participation as content producers and consumers on social media exposes these individuals to a complex set of vulnerabilities and risks, manifesting in different dimensions of their development.

As discussed in the previous section, commercial exploitation emerges as one of the main factors of child and adolescent vulnerability on digital platforms. Unlike traditional artistic work, which in many countries has established legal protections, content production for social media often occurs in a regulatory vacuum (Minnich, 2024). This legal gap leaves children vulnerable to different forms of exploitation, including by their own parents or guardians.

The Hobson family case illustrates the extreme risks of this exploitation: seven children were forced to produce content for a YouTube channel that accumulated over 700,800 followers and 242 million views. To maintain content production, five of the children were deprived of formal education for years. Additionally, they were subjected to physical and psychological abuse when they failed to meet production expectations. Despite the channel generating approximately US$ 300,000 in revenue, the children had no legal right to these earnings (Masterson, 2021).

Impacts on psychosocial development constitute another critical dimension of vulnerabilities faced by child influencers. Constant exposure on social media can significantly affect children's identity formation, values, and social skills. Studies indicate that early exposure to certain types of content can negatively influence self-image perception and self-esteem (Steinsbekk *et al*., 2020).

The pressure for continuous content production, need to respond to fans, and participation in events can generate significant stress levels. Furthermore, there is a risk of early sexualization, with children being placed in inappropriate poses on accounts managed by their parents to increase follower counts. This early objectification can have severe consequences for emotional development and identity formation.

Another critical aspect of the vulnerabilities faced is represented by privacy and data protection. In 2019, YouTube was fined US$ 170 million for violating the Children's Online Privacy Protection Act (COPPA) by collecting personal data from children under 13 without parental consent (Johnson *et al*., 2024). This case illustrates how digital platforms can exploit child vulnerability for commercial purposes, even when specific legal protections exist.

Moreover, the phenomenon of "sharenting," where parents extensively share their children's lives on social media, can result in serious online safety threats. For example, pedophile networks can exploit recommendation algorithms to find content involving children and use comment features and direct messages for illicit purposes (Dorasamy *et al*., 2021).

All this exposure is amplified by sophisticated algorithmic manipulation strategies and persuasive design implemented by digital platforms, aiming to maximize user engagement and time spent. Such strategies include deceptive patterns, such as manufactured time pressure, navigation restrictions, and "bait" mechanisms specifically designed to stimulate longer usage time and engagement. A significant example is the confirmshaming pattern, where users are emotionally manipulated to perform actions they normally wouldn't, such as maintaining an active monetized account or continuing to produce content under time pressure (Mathur *et al*., 2021)

The platforms' persuasive design operates through functionalities structured to modify users' choice set and manipulate information flow. These techniques exploit psychological vulnerabilities to maximize engagement metrics such as time spent, daily visits, and interactions, often contrary to the individual's conscious will. In children's specific case, this manipulation is particularly effective due to their limited self-regulation capacity and understanding of the commercial strategies involved (Albuquerque *et al*., 2024). Thus, manipulation through persuasive design and neurotechnology, using gain and reward methods, can create dependency patterns similar to those observed in gambling (Cruz and Pereira Júnior, 2022).

The impact of this intensive digital exposure on children's cognitive and emotional development is not yet fully understood due to the phenomenon's relative novelty. However, preliminary evidence suggests that pressure for constant performance and early exposure to monetization can compromise healthy personality development and social skills. The combination of these different vulnerability factors creates a complex scenario that demands urgent attention from legislators, digital platforms, and civil society. Effective protection of children and adolescents in the digital environment requires a holistic approach that considers not only legal and technical aspects but also the psychological, social, and neurological impacts of intensive digital exposure.

3. The importance of neurorights in protecting child and teen influencers

The emergence of neurorights as a specific category of legal protection represents a necessary response to the challenges posed by the intersection of neuroscience and digital technologies. According to the Neurorights Initiative (2019), this framework consists of five fundamental rights: the right to identity, understood as the ability to control one's physical and mental integrity; the right to agency, which encompasses freedom of thought and free will in choices; the right to mental privacy, concerning the protection of thoughts against disclosure; the right to fair access to mental enhancement technologies; and the right to protection against algorithmic bias.

As Ana Maria D'Ávila Lopes points out regarding neurorights, "this is a complex theme with clearly innovative contours, whose questions may not find answers in the traditional theory of human rights, which has a clearly ontological character, thus requiring the formulation of new rights" (Lopes, 2023).

In the specific context of childhood and adolescence, the application of these rights takes on particular contours, considering the neurological vulnerability characteristic of this developmental phase. This vulnerability is heightened in the digital

environment where, as seen, behavioral manipulation and persuasive design are employed to maximize engagement and monetization.

In this sense, the management of kidfluencers' authenticity involves tensions between parental control, commercial pressures from advertisers and platforms, and follower expectations, resulting in a distorted digital identity. In other words, there is a strong tendency for child influencer profiles not to mirror these children's real characteristics, but rather an idealized version created by their parents, companies, and followers, which can result in identity fragmentation, where the online persona constructed for commercial purposes overlaps with the natural development of offline personality (Abeele *et al*., 2024). These practices conflict with the neuroright to identity, which seeks to ensure the protection of individual authenticity and freedom of identity construction.

The right to agency is also systematically compromised by the manipulation mechanisms employed by digital platforms. Children's limited capacity for discernment, combined with sophisticated persuasive techniques, can result in compulsive behaviors and choices that are not genuinely autonomous. As previously mentioned, digital platforms commonly operate with "fame gamification" systems, allowing influencers access to engagement metrics such as likes and shares, clearly stimulating the production of more content while diminishing user decision autonomy (Albuquerque *et al*., 2024).

Similarly, the right to keep thoughts protected against involuntary disclosure is constantly challenged by behavioral data collection and analysis practices for commercial purposes. In the case of child influencers, as seen, the issue is aggravated by the phenomenon of sharenting, whereby parents extensively share their children's lives on social media, potentially compromising not only their physical privacy but also their psychic integrity (Walrave *et al*., 2022).

Protection against algorithmic bias assumes special relevance when considering children's cognitive vulnerability to manipulation and algorithmic systems' tendency to amplify stereotypes and problematic behaviors. As evidenced by Feller and Burroughs (2022), there is an emergence of specialized agencies developing sophisticated native advertising strategies targeted at children, frequently exploiting cognitive biases and psychological vulnerabilities.

In summary, the analysis and propagation of neurorights in child and youth protection represents an ethical and legal imperative in face of the challenges posed by the increasing digitalization of children's experiences. The in-depth understanding of these rights and their specific application regarding child digital influencers can contribute to the development of more effective protection mechanisms, ensuring a safer digital environment conducive to the healthy development of children and adolescents.

4. Perspectives for protecting and promoting children and adolescents' neurorights on social media

The effective protection of children and adolescents' neurorights in the digital environment demands a multifaceted approach, involving different actors and levels of intervention. It is essential to analyze the different perspectives and proposals to

ensure the integral protection of these individuals, considering both the existing regulatory frameworks and the new emerging needs of the digital scenario.

In the realm of public policies, the role of the State and regulatory agencies is essential to establish regulatory frameworks appropriate to the specificities of the kidfluencers phenomenon. International experience demonstrates the importance of specific regulation - France, for example, became in 2020 the first country in the world to establish legal rules for digital influencers under 16 years old, following a mobilization initiated in 2018 by the Observatory of Parental and Digital Education (Tomaz *et al*., 2023).

The existing legal frameworks in Brazil already establish important principles for integral protection. The Child and Adolescent Statute (ECA), the Consumer Protection Code (CDC), the Internet Civil Framework, and the General Personal Data Protection Law (LGPD) constitute fundamental mechanisms for understanding Brazilian children's rights in the digital environment. However, the application of these instruments needs to be adapted to the specificities of new forms of digital participation.

Resolution 163 of the National Council for the Rights of Children and Adolescents (Conanda), approved in March 2014, represents a significant effort to regulate advertising directed at children, listing various aspects that may configure abusiveness. Civil society movements view this resolution as an important complement to the microsystem of collective rights protection (Thibau and Rodrigues, 2015).

The advertising sector's self-regulation also plays an important role. In Brazil, the National Council for Advertising Self-Regulation (Conar) code condemns practices commonly used by digital influencers, such as merchandising or indirect advertising using children. Conar's Digital Influencer Advertising Guide establishes guidelines for applying the code's rules to commercial content on social networks, focusing on content generated by users called "influencers" (Tomaz *et al*., 2023).

Recent legislative proposals seek to fill specific regulatory gaps. Bill No. 1222/2024, inspired by actress Larissa Manoela's case, provides for increased penalties for the crime of abuse of incapable persons when committed by parents or guardians, and establishes that young people's corporate participation in their companies must be preceded by the Public Prosecutor's Office's manifestation and contain a revision clause for contract suspension when artists turn 18 (Marzullo, 2024).

Internationally, initiatives such as the KIDS Act, proposed in the United States, seek to regulate deceptive advertising practices targeting children, including manipulative design strategies, influencer marketing, native advertising, and neuromarketing techniques. The proposed legislation aims to prohibit common persuasive design techniques, such as autoplay, push notifications, and manipulative in-app purchase techniques (Feller and Burroughs, 2022).

Child-centered design also emerges as a crucial aspect for neurorights protection. The UK's Children's Code or Age Appropriate Design Code (AADC) establishes important guidelines in this regard, emphasizing how digital products and services design significantly impacts young people's online experiences. The code specifically addresses design techniques intended to influence user behavior, using the term

"nudge" to describe strategies that direct or encourage users to follow developers' preferred paths (Grace *et al*., 2023).

On the other hand, families' role in protecting children and adolescents' neurorights is inescapable and complex. Parents and guardians face significant ethical dilemmas when navigating between parental freedom of expression and children's privacy, especially related to sharenting and commercial exploitation of child influencer activity.

Thus, media education and digital literacy for children and their families constitute essential preventive measures. The United Nations General Comment No. 25 on children's rights in the digital environment, developed after consulting more than 709 children and young people from 28 countries, emphasizes the importance of supporting, promoting, and protecting digital engagement safely and equitably. The document establishes necessary measures to ensure the realization of children's rights concerning the digital environment, assigning specific responsibilities regarding the training of parents and educators on the appropriate use of digital tools (UN, 2021).

Monitoring and evaluating the impacts of implemented protective measures are essential for their continuous improvement. Data collection and analysis on the effectiveness of regulatory frameworks, combined with research on psychological impacts and advancement of neurorights studies, can provide more solid foundations for public policy development.

Therefore, it is necessary to recognize that protecting children and adolescents' neurorights on social media cannot be achieved through isolated measures but requires a coordinated and continuous effort from all involved actors. The successful implementation of the presented protection perspectives depends on the ability to balance the need for protection with the recognition of children and adolescents' autonomy and capacity for expression in the digital environment.

Conclusion

The phenomenon of child digital influencers represents a significant challenge for protecting children's and adolescents' rights in the digital environment. The analysis developed shows that the complexity of this phenomenon demands an approach that transcends traditional legal protections, incorporating the dimension of neurorights as a fundamental framework to ensure the healthy and autonomous development of these individuals.

The conceptual and normative reconfiguration of neurorights in childhood and adolescence proves imperative for addressing the specific vulnerabilities of this age group in face of algorithmic manipulation techniques and persuasive design employed by digital platforms. The protection of identity, agency, mental privacy, and fair access to technologies should be considered priority in the development of public policies and regulatory practices.

The perspectives for protecting and promoting children's and adolescents' neurorights on social networks demand a coordinated effort among different social actors: states, digital platforms, families, and educators need to act in an integrated way to create a safer digital environment conducive to healthy development.

Specific regulation of kidfluencer activity, including labor and patrimonial protection aspects, proves urgent and necessary. However, this regulation must balance rights protection with recognizing children's and adolescents' capacity for expression and social participation in the digital environment.

With this in mind, it would be possible to present a list of issues that could permeate public policies and programs aimed at ensuring due respect for neurorights and the peculiar rights of children and adolescents exposed to the kidfluencer situation, as persons in a peculiar condition of development. The proposal stems from the analysis of researched data, and the questions below can also be structured as topics of a normative diploma that establishes evaluation parameters aimed at the best interest of children and adolescents.

Digital Safety Assessment Checklist for Children and Adolescent

a) Digital Identity and Authenticity: 10 Does the online presence genuinely reflect the child/adolescent's personality and age? 2) Is there external pressure to maintain a certain digital image or persona? 3) Do published contents respect natural identity development? 4-Is there balance between digital exposure and privacy preservation? 5-Does the child/adolescent show discomfort with their online representation?

b) Autonomy and Freedom of Choice: 1-Are decisions about publications primarily made by the child/adolescent? 2-Is there clear understanding about consequences of digital interactions? 3-Do platform "gamification" mechanisms influence compulsive behaviors? 4-Is there freedom to refuse participation in commercial content? 5-Is time dedicated to digital activities self-regulated or externally imposed?

c) Mental Privacy Protection: 1-Are there safeguards against excessive behavioral data collection? 2-Are privacy settings adequately managed? 3-Is there excessive sharing of intimate or vulnerable moments? 4-Is the right to digital forgetting being considered? 5-Is there a protocol for removing sensitive or harmful content?

d) Digital Access and Inclusion: 1-Are available technological tools age-appropriate? 2-Is there adequate technical and educational support? 3-Do platforms used offer effective parental controls? 4-Is there access to updated digital protection resources? 5-Do technologies promote healthy cognitive development?

e) Protection Against Algorithmic Manipulation: 1-Is there monitoring of algorithm-suggested content? 2-Does the child/adolescent recognize native advertising and influencer marketing? 3-Are there tools to filter potentially harmful content? 4-Is there understanding about recommendation systems' functioning? 5-Are strategies used to minimize exposure to addictive content?

f) Family Supervision and Support: 1-Do guardians maintain open dialogue about digital experiences? 2-Is there a structured parental mediation plan? 3-Are usage rules collaboratively established? 4-Is there regular monitoring of online interactions? 5-Is there emotional support to handle digital challenges?

g) Mental Health and Digital Well-being: 1-Are signs of anxiety related to engagement metrics observed? 2-Is there balance between online and offline activities? 3-Is there monitoring of sleep and behavior patterns? 4-Are regular psychological impact assessments conducted? 5-Is professional support available when needed?

h) Digital Literacy and Training: 1-Does the child/adolescent receive education about digital safety? 2-Is there understanding about social engineering risks? 3-Is there training about information verification? 4-Are digital critical thinking skills developed? 5-Is there access to educational resources about digital citizenship?

i) Commercial and Financial Aspects: 1-Is there protection against inappropriate commercial exploitation? 2-Are financial gains adequately managed and protected? 3-Is there transparency in established commercial relationships? 4-Is specialized legal advice available when needed? 5-Are child labor regulations respected?

j) Social and Emotional Development: 1-Do digital interactions promote healthy relationships? 2-Is there space for authentic expression of emotions? 3-Are digital empathy skills developed? 4-Is there balance between online and in-person socialization? 5-Is there support for developing digital resilience?

This evaluative framework should be applied periodically, considering: (a) Age and developmental stage; (b) Family and social context; (c) Purpose of digital platform use; (d) Specific observed impacts; (e) Individual protection needs. Regular evaluation of these aspects allows: (a) Early risk identification; (b) Implementation of preventive measures; (c) Adjustments in protection strategies; (d) Development of digital competencies; (e) Promotion of integral well-being.

Finally, the need to continue developing research and studies about neuropsychological impacts of intensive digital exposure is emphasized, as well as about the effectiveness of implemented protection measures. Only through deep understanding of these aspects will it be possible to develop truly effective policies and practices for comprehensive protection of children's and adolescents' rights in the contemporary digital environment.

References

Abidin, C. (2020). Preschool stars on YouTube: child microcelebrities, commercially viable biographies, and interactions with technology. In: Green, L.; Holloway, D.; Stevenson, K.; Leaver, T.; Haddon, L. (eds.). *Routledge Companion to Digital Media and Children*. London: Routledge, p. 226-234.

Albuquerque, N., Valença, G., Falcão, T.P. (2024). Investigating manipulative design on social media platforms - the case of kidinfluencers. Journal on Interactive Systems, 15(1), 860-874. https://journals-sol.sbc.org.br/index.php/jis/article/view/4271

Andrade, M. (2022). *Child artistic work on digital platforms*: for comprehensive protection of child influencers. Master's Thesis, Federal Rural University of Semi-Arid, Brazil.

CGI.br (2024). *ICT Kids Online Brazil Survey 2023*: Survey on Internet use by children in Brazil. São Paulo: Brazilian Internet Steering Committee. www.cetic.br

Cruz, P.M.M., Cruz, R.V.M. (2024). Neurorights and the challenge of comprehensive protection: investigation of children and adolescent prosumers' activities in the electronic games market. In: Nunes, C.A.R. et al. (eds.). *Proceedings of VIII CIDH* Coimbra 2023 - Volume 4. Campinas/Jundiaí: Brasílica/Edições Brasil.

Cruz, P.M.M., Pereira Júnior, A.J. (2022). Neuroscience and law: social media algorithm interference and its impacts on children and adolescents' human rights. In: Lopes, A.M.D. et al. (eds.). *Neurolaw, neurotechnology and human rights*. Porto Alegre: Livraria do Advogado.

Dorasamy, M., Kaliannan, M., Jambulingam, M., Ramadhan, I., Sivaji, A. (2021). Parents' awareness on online predators: cyber grooming deterrence. *The Qualitative Report,* 26(11), 3683-3723. https://doi.org/10.46743/2160-3715/2021.4914

Feller, G., Burroughs, B. (2022). Branding kidfluencers: regulating content and advertising on YouTube. Television & New Media, 23(6), 575-592. https://doi.org/10.1177/15274764211052882

Geider, J.F. (2021). How to (not) exploit your internet child star: unregulated child labor on YouTube, Instagram and social media. Student Works, 1150. https://scholarship.shu.edu/student_scholarship/1150

Grace, T.D., Abel, C., Salen, K. (2023). *Child-centered design in the digital world*: investigating the implications of the age-appropriate design code for interactive digital media. Proceedings of the 22nd Annual ACM Interaction Design and Children Conference, 289-297. https://doi.org/10.1145/3585088.3589370

Johnson, G., Lin, T., Cooper, J.C, Zhong, L. (2024). *COPPAcalypse?* The YouTube settlement's impact on kids content. SSRN. https://ssrn.com/abstract=4430334

Lopes A.M.D. (2022). Neurotechnology: when technology threatens human dignity. In: LOPES, A.M.D. et al. (eds.). *Neurolaw, neurotechnology and human rights*. Porto Alegre: Livraria do Advogado.

Masterson, M.A. (2021). When play becomes work: child labor laws in the era of "kidfluencers". *University of Pennsylvania Law Review*, 169, 577. https://scholarship.law.upenn.edu/penn_law_review/vol169/iss2/5

Martínez Allué, M., Martín Cárdaba, M.A. (2024). "Kidfluencers": children influencers on YouTube and TikTok and their impact on the child audience. *Visual Review*: International Visual Culture Review, 16(5), 261-270. https://doi.org/10.62161/revvisual.v16.5301

Marzullo, L. (2024). Larissa Manoela Law: CCJ approves bill that increases penalty for parents who commit abuse of incapable. *O Globo*. https://oglobo.globo.com/politica/noticia/2024/06/26/lei-larissa-manoela-ccj-aprova-projeto-que-aumenta-pena-para-pais-que-cometerem-abuso-de-incapaz.ghtml

Mathur, A. Mayer, J., Kshirsagar, M. (2021). What makes a dark pattern... dark?: design attributes, normative considerations, and measurement methods. *CHI Conference on Human Factors in Computing Systems*, 1-13. https://doi.org/10.1145/3411764.344561

Minnich, M. (2024). That's just show business: relying on industrial revolution solutions for a 'kidfluencer' problem. *University of Louisville Law Review*, 62(2). https://ssrn.com/abstract=4789150

Monge Roffarello, A., De Russis, L. (2022). Towards understanding the dark patterns that steal our attention. *ACM Conference on Human Factors in Computing Systems*, 1-7. https://doi.org/10.1145/3491101.3519829

Reardon, J. (2022). New kidfluencers on the block: the need to update California's Coogan Law to ensure adequate protection for child influencers. *Case Western Reserve Law Review*, 73, 165. https://scholarlycommons.law.case.edu/caselrev/vol73/iss1/6

Steinsbekk, S., Wichstrøm, L., Stenseng, F. et al. (2021). The impact of social media use on appearance self-esteem from childhood to adolescence – a 3-wave community study. *Computers in Human Behavior,* 114. https://doi.org/10.1016/j.chb.2020.106528

Thibau, T.C.S.B., Rodrigues, J.P.P. (2016). CONANDA Resolution 163 as a means of protecting children as a collective. *Revista de Direito Brasileira*, 12(5), 68-86. https://doi.org/10.26668/IndexLawJournals/2358-1352/2015.v12i5.2875

Toffler, A. (1980). The Third Wave. New York: Bantam.

Tomaz, R., Guedes, B., Martins, I. (2023). Main challenges for child digital citizenship in a consumer culture in Brazil. *Journalism and Media,* 4, 42-59.

United Nations (UN). (2021). General Comment No. 25 on children's rights in relation to the digital environment. *Committee on the Rights of the Child*. Geneva: UN. https://tinyurl.com/y3dnskx6

Van Den Abeele, E., Hudders, L., Vanwesenbeeck, L. (2024). *Managing authenticity in a kidfluencers' world: a qualitative study with kidfluencers and their parents*. New Media & Society. https://doi.org/10.1177/14614448231222558

Walrave, M., Verswijvel, K., Ouvrein, G., Staes, L., Hallam, L., Hardies, K. (2022). The limits of sharenting: exploring parents' and adolescents' sharenting boundaries through the lens of communication privacy management theory. *Frontiers in Education*, 7. https://doi.org/10.3389/feduc.2022.803393

— 15 —

VERIPOL: legal and technical analysis of the tool for the detection of false reports

MARÍA CONCEPCIÓN RAYÓN BALLESTEROS
Universidad Complutense de Madrid
mcrayon@ucm.es

JOSÉ LUIS GONZÁLEZ ÁVILA
Universidad de la Laguna
jlgavila@ull.edu.es

1. Overview of the VERIPOL tool

VERIPOL is a pioneering artificial intelligence system in the field of public security, implemented by the security forces in Spain, specifically by the National Police,[1] in order to identify false allegations in criminal proceedings. The tool was developed in collaboration with a research team from Cardiff University, taking advantage of advances in natural language processing (NLP) and *machine learning* (Quijano Sanchez; Liberatore; Camacho-Collados; Camacho-Collados, 2018). It was

[1] See the publication of the National Police on the implementation of this tool: https://www.policia.es/_es/comunicacion_prensa_detalle.php?ID=4433&idiomaActual=es

implemented in Spain's National Police Stations in 2018 after a period of testing in several towns.[2]

The core of the VERIPOL tool is based on the use of advanced natural language processing and machine learning techniques.[3] These algorithms allow the system to analyze the linguistic features present in the written complaints submitted by . The artificial intelligence, trained with thousands of previous complaints - both false and truthful-, is able to identify patterns and elements that statistically correlate with the falsity or veracity of the complaint.[4]

In more specific terms, VERIPOL evaluates a series of parameters of the text of the complaint such as narrative structure, amount of detail, internal coherence, type of language used.

The system does not issue a definitive opinion on the falsity of the report, but makes a probabilistic assessment, assigning a percentage of probability that the report is false or true.[5] This result is subsequently used by officers as an indication, which complements other police investigations. The reliability index of the tool is 70 to 90% depending on the source consulted.[6]

2. Purpose and Objectives of the tool

The use of VERIPOL obeys several purposes of public and private interest that contribute to the improvement of the criminal justice system and the efficiency of police resources. Among its main objectives are:

a) Efficiency in resource allocation: Law enforcement is limited by the availability of human and material resources. The investigation of false reports diverts resources from the prosecution of real crimes, which is detrimental both to the functioning of the system and to the victims of real crimes. By detecting potential fraud early, VERIPOL allows for a more efficient allocation of these resources.

b) Protection of the public interest: False complaints constitute an abuse of public reporting mechanisms and a fraud on the judicial system. Their proliferation can cause substantial damage public confidence in the institutions responsible for the administration of justice. VERIPOL, by identifying such fraudulent practices, contributes to safeguarding the integrity of the criminal justice system.

---

[2] It was implemented in 2018 after a testing period in police stations in Malaga and Murcia, specifically the success ratios were 81.58% in Murcia and 84.78% in Malaga. According to published communication from the National Police "the model is able to discern significant differences in the narrative of true and false allegations leading to the best separation between these two classes. From this analysis it can be concluded that true and false reports differ mainly in three main aspects: modus operandi of the aggression, morphosyntax of the report and amount of details. https://www.policia.es/_es/comunicacion_prensa_detalle.php?ID=4433&idiomaActual=es

[3] It uses statistical models trained with sets of true and false allegations to learn the characteristic patterns of each, looking for common word combinations and syntactic structures in false allegations that it contrasts with a database.

[4] It is mainly used to report robberies with violence, intimidation and theft.

[5] It is important to note that VERIPOL does not replace the officers in the investigation, but is a support tool that generates alerts for the police to conduct more detailed investigations.

[6] According to the National Police it has an accuracy rate of over 90%: https://www.policia.es/_es/comunicacion_prensa_detalle.php?ID=4433&idiomaActual=es "VERIPOL: the police tool for detecting false reports", available at this link which quotes police sources, https://www.rtve.es/noticias/20181102/VERIPOL-herramienta-policia-para-detectar-denuncias-falsas/1831344.shtml According to published several newspapers national newspapers: https://www.eldiario.es/navarra/policia- nacional-introduce-sistemas- denuncias_1_1658984.html.

c) Fraud deterrence: The mere existence of a tool such as VERIPOL can have a deterrent effect on potential criminals who, being aware that their reports will be subject to a thorough and objective analysis, may refrain from filing false reports.

Although VERIPOL is a tool with a high rate of accuracy, it is essential to emphasize that its use is subject to legal limitations and guarantees that seek to protect the fundamental rights of complainants. In accordance with the principles of the rule of law, the use of artificial intelligence tools in the administration of justice must be carefully regulated to avoid arbitrariness. In this sense, VERIPOL does not replace in any case the human assessment of the facts or the traditional police investigation. Its function is merely auxiliary, providing an indication that must be corroborated by other evidence. The police remain responsible for investigating the facts and ensuring that the procedural rights of complainants, such as the right effective judicial protection, the presumption of truthfulness and the right to a fair trial, are respected.

We will then analyze the technical and legal aspects of the tool to finally reach the corresponding conclusions regarding the operation of VERIPOL.

3. Technical aspects to consider in VERIPOL

VERIPOL is based on the application of state-of-the-art methodologies in natural language processing (NLP) and *machine learning* (Martínez; Delgado; Pérez, 2018),[7] which enable automated analysis of the textual content of the statements made.

Among the main technological resources used by VERIPOL, the following stand out:

a) Natural Language Processing (NLP): This technique allows us to unravel the structure and linguistic style of the allegations, identifying characteristic idiomatic patterns (Camacho-Collados et al., 2016)[8] that tend to differentiate truthful from fraudulent accounts.

b) Machine learning algorithms: VERIPOL has been trained by analyzing thousands of reports (Ramírez et al., 2017),[9] both genuine and false, which allows it to discern regularities in fraudulent reports, such as the scarcity of details, narrative inconsistency or the use of expressions and terms that tend to betray falsehood.

c) Pattern analysis and narrative complexity: VERIPOL's algorithms are designed to identify excessive simplicity or lack of cohesion in the narrative of the facts, recurring characteristics in fraudulent allegations.

3.1. Algorithm and underlying technology

The machine learning and natural language processing (NLP) algorithms implemented in VERIPOL represent sophisticated mechanisms designed for the thorough

[7] "The incorporation of machine learning techniques has allowed VERIPOL to accurately identify linguistic cues typical of false reports, overcoming traditional manual analysis methods" (Martínez et al., 2018).

[8] "The use of natural language processing to detect fraudulent patterns in complaints has proven to be an effective tool in public safety" (Camacho-Collados et al., 2016).

[9] "The evaluation of VERIPOL in real contexts has shown an accuracy of more than 90% in the detection of fraudulent allegations, especially in robbery-related statements" (Ramírez et al., 2017).

scrutiny of text, with the purpose of detecting falsehoods (Méndez et al., 2017).[10] in formal statementsSuch algorithms, fed by extensive volumes of data - including authentic and fraudulent claims - allow the identification of linguistic and stylistic patterns that escape conscious human perception. By analyzing the deep structure, content and time evolution of texts, these algorithms are able to more accurately discern indications of deception in written communications.

3.1.1. Language as a Reflection of Deception: Language, as a manifestation of human thought, carries complex patterns that sometimes reveal psychological states or hidden intentions. Lie detection through linguistic analysis is based on the premise that certain stylistic features emerge when a person tries to hide the truth. Research, such as that conducted by Brigham Young University (BYU), has shown that, in digital interactions, individuals who lie tend to take longer to respond, revise their writing more frequently, and generate shorter, more carefully controlled responses. VERIPOL's algorithms integrate these types of observations and recognize them systematically.

3.1.2. Psychological Heuristics in VERIPO: VERIPOL incorporates heuristics based on behavioral analysis derived from psychological studies on deception. Among the clues that its algorithms can identify are the scarcity of details, the lack of narrative coherence, the use of vague phrases or the repeated revision of the text, all considered as potential markers of falsehood. These heuristics suggest a correspondence with certain psychological criteria employed in lie detection, linking linguistic analysis with widely accepted principles in cognitive psychology. However, it is crucial to emphasize that these techniques do not provide absolute determinism; rather, conclusions derived from VERIPOL are based on statistically calculated probabilities.

3.1.3 Continnuos Learning Capability: Currently, VERIPOL operates under a supervised learning model, but raises the possibility of expanding to a continuous (López et al., 2018)[11] or unsupervised learning system, which would allow for a greater adaptability over time. While its current capability is based on predefined data sets, its integration with new data - for example, judicial rulings that confirm or refute the veracity of the allegations - would increase its predictive accuracy. This would imply a constant feedback of the system, allowing it to evolve in tune with judicial verdicts, further refining its ability to detect fraudulent allegations.

3.1.4. The Digital Deception Challenge: The realm of digital communications presents an added challenge in detecting deception, given the lack of nonverbal cues that are normally used to identify lies. As noted by expert Tom Meservy, the digital environment is conducive to deception because of the ease with which people can conceal their identity and construct narratives that, on the surface, appear credible. However, VERIPOL's algorithms mitigate this challenge by focusing on linguistic patterns (García et al., 2018)[12] and the temporal behavior of written responses, obtain-

[10] "VERIPOL represents a significant advance in the automation of processes within law enforcement, enabling faster and more efficient assessment of potentially false reports" (Méndez et al., 2017).

[11] "Continuous learning allow tools such as VERIPOL to improve their accuracy over time by incorporating new allegations and behavioral patterns" (López et al., 2018).

[12] "VERIPOL has been shown to be effective in detecting common linguistic patterns in fraudulent complaints, such as excessive use of irrelevant details and narrative simplification" (García et al., 2018).

ing a higher falsehood detection rate than conventional methods, which in previous studies barely reach 54% accuracy.

3.2. Natural Language Processing (NLP)

VERIPOL relies on the most advanced techniques of Natural Language Processing (NLP) and machine learning, in order to identify linguistic and stylistic patterns that could reveal the falsity of a complaint.

3.2.1. VERIPOL System Technical Components: The VERIPOL system is built on an architecture composed of several key elements that interact with each other, especially Natural Language Processing (NLP). In this area, a meticulous analysis of the text of the reports is carried out, in order to extract features and characteristics that may reveal inconsistencies or structures typically associated with fraudulent reports. The process focuses on identifying key words, unusual syntactic structures and narrative patterns. These elements are then processed and classified using labeling techniques, which assign grammatical roles to different parts of speech to assess their relevance in determining the veracity of the claim.

3.2.2. Use of the Natural Language Toolkit (NLTK) in VERIPOL: One of the fundamental pillars of linguistic processing in VERIPOL is the use of the well-known Python Natural Language Toolkit (NLTK) library (Bird, Klein & Loper, 2009).[13] NLTK has established itself as one of the most advanced and versatile tools for natural language processing, offering an extensive repertoire of functions and resources designed for textual data analysis. In VERIPOL, NLTK plays a crucial role in several processing stages, among which the following stand out:

a) Tokenization: This technique allows the text to be broken down into its smaller components, whether words or phrases, for more detailed analysis.

b) Parts of Speech Tagging (POS tagging): Through this technique, NLTK identifies the grammatical categories (nouns, verbs, adjectives, etc.) within the text, providing a structural view of the text.

c) Lemmatization and stemming: NLTK reduces words to their base or root form, facilitating uniformity in the analysis and eliminating irrelevant morphological variations.

d) Feature extraction: The system extracts key stylistic and linguistic properties of the text that are indicative of its authentic or false nature.

e) Syntactic and semantic analysis: NLTK allows to unravel the underlying grammatical structure of the text, as well as to understand its overall meaning, both essential aspects in the identification of fraudulent patterns.

3.2.3. Integration with Machine Learning: VERIPOL not only limits itself to analyzing language from a linguistic perspective, but also integrates the results of this analysis with machine learning techniques. Here, NLTK plays a central role by preprocessing the textual data, preparing the texts for the learning algorithms to work with. Through transformation to feature vectors, the texts are converted into numeri-

[13] "NLTK offers a wide range of tools for natural language processing, facilitating tokenization, lemmatization, and parsing" (Bird, Klein & Loper, 2009).

cal representations that machine learning models (Delgado et al., 2017)[14] can interpret to make predictions

The combination of PLN and machine learning enables VERIPOL to detect indications of fraudulent claims with high accuracy. The system learns from large volumes of historical data, both true and fraudulent reports, and as it processes more information, it is able to refine its predictive models.

3.2.4. Complement with Weka: Another key component of the system is the use of Weka (Witten & Frank, 2005),[15] a powerful machine learning tool. NLTK and Weka work in tandem: while NLTK handles the initial feature extraction and labeling, Weka is used for the more advanced classification and prediction tasks. The interoperability between these two platforms allows data processed by NLTK to be exported in a Weka-compatible format, thus optimizing the training and evaluation of learning models.

3.2.5. NLTK, Leading tool for Natural Language Processing: NLTK has positioned itself as a leading platform for the development of Python programs that work with human language data. Through its user-friendly interface and its access to more than 50 corpora and lexical resources, such as WordNet, NLTK allows a deep analysis of language in its various dimensions. Its multiple functions include classification, tokenization, stemming, tagging, syntactic analysis and semantic reasoning, making this library a reference in the field of computational linguistics.

In addition, NLTK integrates easily with other industrial-strength text processing tools and libraries, making it an ideal choice for both academic and commercial applications. Its flexibility and the support of an active developer community have enabled NLTK to become a robust open source platform widely used in natural language processing projects around the world.

In the context of VERIPOL, NLTK not only provides the fundamental resources for linguistic analysis, but is also instrumental in the development of algorithms that enable continuous refinement of the system. As it is updated with new data and court rulings, the system continues to learn and improve its ability to detect falsehoods with greater accuracy.

3.3. Tokenization in NLTK: The text segmentation process

At its core, tokenization consists of breaking down the text into minimal units of analysis, called "tokens", which can be words, phrases or characters depending on the desired level of granularity. In VERIPOL, tokenization is mainly performed at the word level, since the structure and choice of words in a specific context can reveal key clues as to the veracity or falsity of an allegation.

In NLTK, this process is performed using a set of tools optimized for various languages and text structures. The steps include lexical segmentation, where continuous text is separated into words using the WordPunctTokenizer algorithm, and linguistic exception handling, which adjusts the processing of complex terms such

[14] "VERIPOL uses machine learning models trained on historical data to identify fraudulent claims using linguistic and syntactic analysis" (Delgado et al., 2017).

[15] "Weka enables the implementation of machine learning algorithms and classification techniques on textual data" (Witten & Frank, 2005).

as contractions or proper nouns. In addition, sentence tokenization, performed using the Punkt tokenizer, allows the text to be broken down into sentences, facilitating the analysis of narrative coherence, which is crucial for detecting inconsistencies in fraudulent claims.

After tokenization, the data is prepared for machine learning algorithms by vectorization. The tokens are converted into feature vectors, using techniques such as Bag of Words or TF-IDF, which assign numerical representations to the terms. This numerical data is exported to Weka, where it is processed in a compatible format such as ARFF, allowing predictive models (Pedregosa et al., 2011)[16] to be trained to identify patterns of falsity

This collaborative cycle between NLTK and Weka is not merely sequential; both systems feed back to each other, continuously adjusting the tokenization and classification process. Through this synergy, VERIPOL achieves high levels of accuracy in detecting fraudulent claims, combining detailed linguistic analysis with the power of machine learning algorithms.

3.3.1.. Part-of-Speech Labeling Process with NLTK: POS tagging refers to the assignment of grammatical categories to each word in a text, such as noun, verb, adjective, among others. This process is fundamental in VERIPOL, as it allows unraveling the underlying grammatical structure of a complaint and discovering patterns indicative of deception.

NLTK, a key tool in this process, employs statistical model-based taggers, such as Treebank trained on Spanish corpora, which assign grammatical tags based on context. This analysis reveals that fraudulent claims tend to present simplified structures, with limited use of adjectives and a predominance of verbs in simple forms.

In addition to its accuracy, NLTK employs probabilistic models, such as those based on Hidden Markov Models (HMM), which handle linguistic ambiguity by assessing the probability that a word belongs to aspecific grammatical category. This probabilistic approach greatly improves the accuracy of falsehood detection.

Integration with Weka converts the labeled data into numerical vectors that feed the classification algorithms (Hernandez et al., 2019).[17] The encoded grammatical categories are processed by Weka to determine whether the detected distributions are consistent with those observed in fraudulent claims, using models such as decision trees or support vector machines (SVM).

This labeling process, moreover, is not static. As Weka classifies allegations, the system feeds back to NLTK, improving the accuracy of part-of-speech labeling and refining falsity predictions. Thus, VERIPOL continually refines its ability to detect syntactic inconsistencies, narrative simplicity, and subtle grammatical manipulations, making linguistic analysis a critical tool for identifying fraudulent claims with greater rigor and sophistication.

[16] "Tools such as Weka and scikit-learn allow the use of predictive algorithms to analyze large volumes of data, improving the accuracy of truthfulness models" (Pedregosa et al., 2011).

[17] "VERIPOL uses a combination of natural language processing and classification to assist police forces detecting false reports" (Hernandez et al., 2019).

3.3.2. Lemmatization and Stemming in the Context of Linguistic Processing: Both lemmatization and stemming[18] seek to simplify words, but differ in their approach. While stemming uses heuristic rules to trim words to their root, lemmatization reduces words to their base form (lemma), considering their grammatical category and context. In VERIPOL, both methods are employed strategically according to the needs of the analysis.

a) Stemming: Reduction to the root: Stemming, implemented in NLTK through algorithms such as Porter Stemmer or Snowball Stemmer, removes suffixes to reduce words to their roots. Although this approach speeds up text processing, the roots obtained are not always real words, which limits their semantic understanding. Nevertheless, in VERIPOL, it is useful for detecting repetitive patterns in denunciations, especially in situations where efficiency is more important than lexical accuracy. By applying stemming, morphological variations such as "steal", "stolen" and "steal" are treated as a single entity, facilitating the identification of simplified stylistic structures characteristic of fraudulent allegations.

b) Lemmatization: Contextualized Analysis: Lemmatization, on the other hand, provides a more accurate analysis by reducing words to their base form, taking context and grammar into account. Using NLTK's WordNet-based lemmatizer, VERIPOL disambiguates words with multiple forms or meanings, which is crucial for detecting fraudulent narratives. For example, a conjugated verb such as "running" is reduced to "running", preserving the original meaning and preventing different forms of the word from fragmenting the pattern analysis.

c) Weka Integration: Preprocessing and Machine Learning: Both lemmatization and stemming prepare the text for processing by the machine learning algorithms in Weka. Words reduced to their base or root form are converted into feature vectors reflecting their relevance and frequency in the text. Tools such as TF-IDF (Term Frequency-Inverse Document Frequency) allow the importance of each term to be measured, helping the Weka algorithms, such as decision trees or SVMs (support vector machines), to classify the allegations as true or false.

d) Impact on False Claims Detection: The joint application of lemmatization and stemming in VERIPOL optimizes the detection of fraudulent claims. By reducing words to their essential forms, linguistic analysis is simplified, allowing the system to identify stylistic patterns that would otherwise go unnoticed. Key benefits include:

- **Language standardization**: Facilitates uniform treatment of different morphological forms of a word, identifying recurring patterns indicative of fraud.

- **Machine learning optimization**: Weka algorithms can process normalized texts more efficiently, improving the detection of relationships between words and contexts.

- **Improved accuracy**: By eliminating the "noise" caused by morphological variations, the system focuses on the linguistic aspects crucial for detecting falsehoods.

[18] Stemming is a method of reducing a word to its root or stem. There are some stemming algorithms that help in information retrieval systems. Stemming increases recall which is a measure of the number of documents that can be found with a query.

3.4. Contextualization of Feature Extraction in VERIPOL

The primary objective of feature extraction in VERIPOL is to identify key linguistic elements that may reveal indications of falsity in the allegations. To do this, the system decomposes the text into essential components, or features, ranging from grammatical structure and narrative style to the semantics used. These features are transformed into vectors that the classification algorithms analyze to assess the veracity of the allegations.

The approach taken by VERIPOL focuses on capturing the underlying properties of language that might go unnoticed in a cursory analysis, such as the frequency of certain terms, syntactic complexity and semantic coherence. By combining NLTK tools and Weka algorithms, the system detects patterns that are often associated with fraudulent narratives, thus offering a substantial advantage in identifying falsehoods.

3.4.1. Feature Extraction Processes with NLTK: The feature extraction process in VERIPOL begins with the exhaustive use of NLTK capabilities, a key library in natural language processing (NLP). This decomposition includes several analytical levels:

a) Tokenization and Segmentation: The text is divided into smaller units, such as words or phrases, allowing a detailed analysis of the underlying grammatical structures.

b) Parts of Speech Tagging (POS Tagging): Grammatical tags are assigned to words, allowing the identification of stylistic patterns associated with fraud, such as the excessive use of adjectives or impersonal forms.

c) Lemmatization and Stemming: These processes reduce words to their base form, avoiding redundancies when analyzing morphological variations and facilitating a more coherent analysis.

d) N-gram extraction: VERIPOL also uses the n-gram technique to analyze specific word combinations, allowing the detection of recurrent speech patterns (Cabrera et al., 2019)[19] or stylistic anomalies.

e) Frequency of Terms and Lexical-Semantic Analysis: VERIPOL analyzes the frequency of use of certain terms, which may reveal attempts to artificially emphasize key aspects of the story.

3.4.2. Vectorization and Transformation with Weka: After extracting the relevant linguistic features with NLTK, these are transformed into feature vectors for processing by the machine learning algorithms in Weka. This vectorization process converts the textual data into numerical representations that the algorithms can interpret. One of the key approaches is the TF-IDF transformation, which measures the relative importance of words in the complaint set.

Subsequently, the vectors are normalized and filtered to remove irrelevant features, thus optimizing the performance of automatic classification models such as decision trees, SVM or Bayesian models.

[19] "The use of natural language processing (NLP) tools in VERIPOL has been instrumental in identifying speech patterns that are not easily detectable by humans" (Cabrera et al., 2019).

3.4.3. Process Impact and Benefits: VERIPOL's efficiency in feature extraction has a decisive impact on false report detection. By combining linguistic, lexical and syntactic features, the system detects subtle patterns that would not be perceptible to a human observer. In addition, continuous feedback between NLTK and Weka allows for progressive learning, refining the detection models.

Finally, the ability to scale the analysis of large volumes of data in an automated manner makes VERIPOL a highly effective tool, enabling law enforcement to address large-scale fraud with precision and efficiency.

3.5. General Context of Syntactic and Semantic Analysis in VERIPOL

Syntactic analysis in this context consists of interpreting the grammatical structure of the text, i.e., how words and phrases are organized to form meaningful sentences. This phase is essential to unravel the structural skeleton of the language in the allegations, identifying grammatical patterns that can differentiate a sincere statement from a fraudulent one.

Semantic analysis, on the other hand, seeks to understand the meaning of these structures by evaluating how words relate to each other to form a coherent narrative. In the case of VERIPOL, semantic analysis helps determine whether statements made in a complaint present logical inconsistencies or indications of fabrication.

3.5.1. Syntactic Analysis with NLTK: Within VERIPOL text processing, parsing is facilitated by the capabilities of NLTK, which allows decomposing the text into its underlying grammatical structure.

a) Syntactic Parser: NLTK includes parsers that build syntactic trees for text sentences. These trees graphically represent how words and phrases are structured within a sentence, providing a detailed view of the grammatical relationships between text components. The syntactic parser examines how nouns, verbs, adjectives and other parts of speech are linked according to grammatical rules.

- Noun and Verb Phrase Analysis: VERIPOL uses this analysis to identify recurring patterns in the organization of phrases. For example, fraudulent allegations may be characterized by an excessive use of simple or fragmented sentences, which may be indicative of an effort to construct a narrative that appears credible but lacks complexity typical of true stories.

b) Named Entity Recognition (NER): An essential part of parsing in VERIPOL is named entity recognition, which allows identifying names of people, places, dates and other entities appearing in text. NLTK offers modules for the recognition of these entities, which is fundamental in the analysis of allegations, since named entities can reveal patterns of manipulation or overload of superfluous details, typical of false statements.

c) Chinking and Chunking: NLTK employs chunking techniques, which group words into broader phrases or segments based on grammatical patterns. In turn, the reverse process, known as chinking, excludes certain elements to focus the analysis on relevant structures. In VERIPOL, these techniques allow a more precise segmentation of sentences, helping to identify suspicious word combinations, such as overly simple narratives or those lacking a clear narrative context.

3.5.2. Semantic Analysis with NLTK: Semantic analysis goes beyond mere structure, focusing on the meaning underlying sentences and how concepts are interrelated to create a coherent discourse.

a) Lemmatization and Meaning Normalization: NLTK is also used to perform lemmatization, that is, to reduce words to their base form. This technique is useful for unifying terms that may have morphological variations, allowing for a more homogeneous analysis of meaning. By reducing words to their roots, the system can identify semantic discrepancies where a complainant uses words in an inconsistent or unusual way.

b) Semantic N-Gram Analysis: Semantic n-grams are another technique employed by VERIPOL to capture sequences of words and analyze their semantic relationship within the text. Through this technique, NLTK evaluates associations between nearby words, allowing to detect inconsistencies in the narrative or unnatural uses of language. For example, a false report may include sentences that, although grammatically correct, do not have a logical semantic relationship, which is an indication of deception.

c) WordNet and Semantic Relationships: NLTK integrates WordNet, a lexical database that groups words into semantic categories such as synonyms, antonyms and hyponyms. In VERIPOL, this tool is used to identify semantic relationships that help unravel the hidden meaning in the allegations. Incorrect use of synonyms, the presence of antonyms in inappropriate contexts, or a lack of lexical variation may indicate that the reporter is fabricating a story.

- Hyponymy and Hyperonymy: The relationship between more general terms (hyperonyms) and more specific terms (hyponyms) is also analyzed to detect inconsistencies in the level of detail. For example, a false report (Navarro et al., 2019)[20] may make excessive use of hyperonyms to avoid specifying details that could be falsifiable.

3.6. Weka Integration: Semantic Transformation in Predictive Models

Once NLTK has performed the syntactic and semantic analysis, these features are transformed into data that can be processed by Weka for classification into predictive models (Rodriguez et al., 2018).[21] In this phase, we seek to identify patterns that correlate syntactic structures and semantic meanings with allegations previously labeled as true or false.

a) Transformation of Syntactic and Semantic Features into Vectors: Grammatical and semantic relations extracted from the text are converted into feature vectors. These vectors include information about sentence structure (such as the relationship between nouns and verbs), semantic coherence (relations between words) and the use of named entities, among other aspects. Weka uses these vectors as inputs for its classification model.

[20] "VERIPOL has been shown to significantly reduce the time and costs associated with investigating false allegations, increasing the operational efficiency of police forces" (Navarro et al., 2019).

[21] "Predictive models implemented in VERIPOL allow to predict with high accuracy the probability of falsehood in complaints from linguistic and semantic patterns" (Rodriguez et al., 2018).

b) Predictive Model Training: Syntactic and semantic features are integrated into Weka's machine learning algorithms, such as support vector machines or decision trees, which look for patterns in fraudulent claims. This approach allows VERIPOL to improve its accuracy as it learns to detect narratives that are syntactically correct, but semantically inconsistent.

c) Narrative Coherence Assessment: One of the most important aspects of integration with Weka is the ability to assess overall narrative coherence. Semantic analysis offers a more holistic view of the discourse, allowing Weka to identify inconsistencies in the narrative flow that could be indicative of a fabricated narrative. This is particularly useful in the analysis of false allegations, where whistleblowers often attempt to construct stories that are credible at a superficial level, but lack internal coherence when analyzed for their deeper meaning.

3.7. Impact of Syntactic and Semantic Analysis on VERIPOL Efficacy

Syntactic and semantic analysis, facilitated by NLTK and Weka, gives VERIPOL a unique ability penetrate into the deepest layers of language and uncover inconsistencies that would otherwise go unnoticed. This combination of techniques allows:

a) Detection of Fabricated Narratives: Syntactic and semantic analysis allows VERIPOL to identify stylistic and narrative patterns typical of false statements. Fraudulent narratives tend to have a less complex structure and lack semantic coherence, factors that can be effectively detected using this technology.

b) Continuous Improvement in Classification: As the system processes more complaints, the feedback obtained allows the Weka-based machine learning models to be refined, which further refines VERIPOL's ability to detect fraud with ever-increasing accuracy.

c) Reduction of False Positives: Semantic analysis makes it possible to reduce the rate of false positives, i.e. truthful statements that initially could have been classified as false due to syntactic or grammatical errors. By evaluating both structure and meaning, VERIPOL ensures a more accurate classification

3.8. Machine learning algorithms

VERIPOL employs a series of advanced machine learning algorithms for the purpose of detecting fraudulent claims through language analysis. These algorithms are trained to identify linguistic and stylistic patterns that indicate the falsity of a statement. In the following, I describe the most prominent algorithms, applying educated and technical language, according to the documentation provided and my specialized knowledge:

a) Decision Trees: The Decision Trees algorithm is based on a hierarchical structure where successive decisions, represented in nodes, guide the analysis of the system. In the context of VERIPOL, each node represents a question about a linguistic feature (such as term frequency or grammatical structure), and the branches of the tree represent the possible answers. This structure makes it easy for the system to classify reports as true or false. The strength of decision trees lies in their ability to decompose a complex problem into a set of binary decisions, which simplifies the analysis of large volumes of textual data.

b) Support Vector Machines (SVM): Support Vector Machines (SVM) are another mainstay in VERIPOL's arsenal of algorithms. This method uses geometry to draw a hyperplane that separates different kinds of allegations (true and false) in a multidimensional space. The SVM approach is particularly effective when working with data that is not linearly separable, as this algorithm can transform the data into a higher dimensionality space using kernel functions, maximizing the separation between classes. The accuracy of SVMs in VERIPOL lies in their ability to handle complex linguistic features, such as semantic analysis and narrative inconsistency detection.

c) Bayesian Models: The Bayesian classifier, based on Bayes' theorem, is used to estimate the probability that a report is false based on the features extracted from the text. This algorithm assumes that the features are independent of each other (conditional independence hypothesis) and calculates the probability that an allegation belongs to a certain class (false or true) based on the previous data. In VERIPOL, the Bayesian classifier has an outstanding predictive value, as it analyzes the combination of linguistic elements such as parts of speech and grammatical patterns, attributing to each one a probabilistic weighting that influences the final decision.

d) Cross Validation and Evaluation Metrics: A crucial aspect in the implementation of these algorithms is cross-validation (Kohavi, 1995),[22] which allows VERIPOL to measure the effectiveness of its models. This technique divides the dataset into chunks, one of which is used to evaluate the model's performance while the rest is used to train it. This ensures that the model generalizes well to previously unseen data. Metrics used to assess accuracy include:

- **Accuracy:** Measures the proportion of true positives in the model predictions.

- **Recall:** Evaluates the model's ability to identify all positive cases

- **F1-Score:** The harmonic mean between precision and recall, providing a balanced measure of performance.

- **ROC curve and AUC:** These evaluate the model's ability to distinguish between classes, with the Area Under the Curve (AUC) being a key metric for assessing its overall effectiveness.

e) Optimization and Continuous Updating: VERIPOL not only benefits from the predictive power of these algorithms, but implements continuous improvement processes through feedback with new data and human reviews. The models are periodically retrained, adapting to changes in reporting patterns and adjusting the hyperparameters of the algorithms to improve their performance. Also, the incorporation of new features, derived from advances in linguistic analysis, strengthens the predictive capability of the system.

f) Integration with Apache OpenNLP and Weka: As for the technical infrastructure, VERIPOL integrates Apache OpenNLP (Apache OpenNLP, 2010).[23] to perform natural language preprocessing (NLP) and extract semantic and syntactic features from the text. Subsequently, the classification algorithms, such as SVM and Bayesian

[22] "Cross-validation is essential in training predictive models to avoid overfitting and to ensure generalization to new data sets"(Kohavi, 1995).

[23] "The use of Apache OpenNLP enables advanced natural language processing through techniques such as tokenization, part-of-speech tagging, and named entity recognition" (Apache OpenNLP, 2010).

models, are executed in Weka, a machine learning platform that facilitates the vectorization of textual data and the efficient processing of large volumes of data.

4. Legal aspects to be considered in VERIPOL

The use of artificial intelligence (AI) in the legal field raises important questions about fundamental rights, due process and legal guarantees. In this context, VERIPOL as a support tool[24] for the detection of false allegations, introduces a framework of regulation and action that must be subject to basic principles of criminal law, criminal procedure, and of course, constitutional law, in order to guarantee legality and justice in its application.

The main legal aspects surrounding the use of VERIPOL are developed below:

4.1. Principle of legality

The principle of legality, which is enshrined in Article 9.3 of the Spanish Constitution, establishes that all actions of the Administration, including State security forces, must be subject to the law. In the context of VERIPOL, the use of this tool must have an adequate regulatory basis that regulates its implementation and limits. This ensures that the tool not used arbitrarily, but within the legally established parameters.

VERIPOL cannot replace the police investigation or the evidentiary analysis carried out by judges or prosecutors. Its use must be complementary and limited to acting as an additional indication, which must be corroborated or disproved by the evidence obtained within the framework of the formal investigation

4.2. Presumption of innocence

One of the most relevant aspects from a legal point of view is the presumption of innocence, a fundamental right enshrined in Article 24.2 of the Spanish Constitution. This basic principle in criminal law guarantees that everyone is innocent until proven guilty. VERIPOL, by issuing a probability of falsehood in the complaint, should not in any case violate this right, because it can not be considered that a complaint is false simply by the result that the tool yields.

In this sense, the information provided by VERIPOL cannot be used as conclusive evidence of falsity, but must be seen only as one more element in the overall assessment of the facts. The principle of presumption of innocence must be respected at all times, both for the complainant and the accused, ensuring that any sanction or legal consequence arising from the complaint is based on solid evidence obtained throughout the criminal process.

4.3. Right to a fair trial and due process

The use of technology such as VERIPOL in the judicial and police sphere poses important challenges with respect to the right to a fair trial, also regulated in Article

[24] It is important to note that VERIPOL does not replace officers in the investigation, but is a support tool that generates alerts for the police to conduct more detailed investigations.

24 of the Spanish Constitution. Any citizen has the right to have his or her complaint investigated impartially and with all procedural guarantees, without an automated tool being able to unduly influence the perception of its credibility.

In addition, it is imperative that transparency in the use of the tool is maintained and that both complainants and defendants are aware of how this technology is being used in their case. This is particularly relevant if VERIPOL results are used to make decisions that affect the course of an investigation. Information obtained through AI algorithms should not be used in an opaque manner or without the possibility of review and rebuttal, as this could violate the principle of the right to defense.

4.4. Right to the protection of personal data and privacy

As a whistleblower analysis tool, VERIPOL processes a large volume of personal data, including sensitive data related to crime reports. In this sense, it must ensure compliance with the data protection regulations in force, mainly the General Data Protection Regulation (GDPR) of the European Union and the Organic Law on the Protection of Personal Data and Guarantee of Digital Rights (LOPDGDD) in Spain.

The use of data by VERIPOL must be duly justified, limited to the purpose of fraud detection and, in any case, comply with the principles of data minimization and confidentiality. The responsibility for the processing of this data lies with the authorities using the tool, who must ensure that there is no violation of the right to privacy of complainants or persons involved in the cases analyzed.

4.5. Dissuasive impact on criminal law: false reporting

VERIPOL's main objective is the detection of false allegations,[25] as the Spanish Penal Code punishes those who, with knowledge of their falsehood or reckless disregard for the truth, accuse another person of acts which, if true, would constitute a criminal offense, thereby triggering the initiation of proceedings.

VERIPOL can play a preventive role in deterring potential false whistleblowers. However, it is necessary to remember that for a report to be considered false in legal terms, it is essential to prove not only the falsity of the reported fact, but also the malicious intent of the complainant. This means that it must be proven that the whistleblower was aware of the falsity of the fact or acted with a clear disregard for the truth.

VERIPOL's automatic detection of inconsistencies or typical patterns of false allegations does not replace the thorough examination of the complainant's intention by judicial agents. Thus, the tool only suggests a possible indication of falsehood, which must be corroborated in the judicial process with due evidentiary rigor.

4.6. The principle of proportionality

The principle of proportionality, enshrined in Spanish criminal and procedural law, is one of the fundamental principles in the actions of public authorities. This

[25] False accusation constitutes a crime typified in the Penal Code in Title XX "Crimes against the Administration of Justice", Chapter V "False accusation and accusation and simulation of crimes" and more specifically in Art. 456.

principle requires that any measure or tool used by the State, including artificial intelligence technologies such as VERIPOL, must be appropriate, necessary and proportional to achieve the legitimate objective it pursues.

In the context of VERIPOL, this means that the use of the tool must be proportional in relation to the objective it pursues, which is the identification of false allegations. Proportionality is assessed at three levels:

a) Adequacy: VERIPOL must be adequate to detect false reports without affecting fundamental rights, and must have demonstrated its effectiveness in meeting this objective.

b) Necessity: The use of the tool must be necessary, i.e., there must not be another less intrusive or restrictive measure that achieves the same result with the same effectiveness.

c) Proportionality in the strict sense: The benefits of using VERIPOL must outweigh the harm it may cause. This implies that, when implementing the tool, it must be assessed whether the possible impact on individual rights, such as the presumption of truthfulness of the whistleblower, is justified against the public interest of preventing fraud in the criminal justice system.

4.7. Application of general principles in relation to the burden of proof

A key aspect of VERIPOL's legality is related to the distribution of the burden of proof in criminal law. Under the principles of the Spanish criminal system, the burden of proof is on the accuser, who must prove the guilt of the accused beyond a reasonable doubt.

VERIPOL, in its role of analyzing the veracity of complaints, can in no case reverse this burden of proof or create a presumption of guilt for the complainant, which would be contrary to the fundamental principles of criminal law. The detection of patterns suggesting falsity in a complaint cannot, by itself, constitute conclusive proof that the complainant has committed a crime of false reporting. Consequently, the results obtained by VERIPOL should be handled as evidence that complements the investigation, and not as the main or only evidence on which to base an accusation or file the complaint

4.8. Judicial evaluation of the evidence provided by VERIPOL

The use of VERIPOL introduces a relevant legal question: what weight should the results of an artificial intelligence tool have in criminal proceedings? In this regard, it is essential to remember that, under Spanish procedural law, judges have the power to freely assess evidence (principle of free assessment of evidence). This means that the result produced by VERIPOL, although it can be used as one more element, must be weighed together with other evidence collected during the police investigation.

Therefore, the judge cannot base a decision exclusively on the VERIPOL report, without it supported by additional objective or testimonial evidence that corroborates the alleged falsity of the complaint. The evidence must be evaluated in its totality and in context, avoiding giving technology a higher decisive value than traditional means of proof, such as statements, expert, or documentary evidence.

4.9. Right to transparency and to the motivation of judicial decisions

In relation to judicial transparency, complainants who are affected by VERIPOL results have the right to know how the tool has been used in their specific case. This right derives from the general principle of motivation of judicial decisions.[26] Any decision taken on the basis of the evidence provided by VERIPOL must be sufficiently reasoned and explain in a clear and understandable way how the judge has assessed the results obtained from the AI tool.

In addition, under the right to defense and the right to due process,[27] the complainant has the right to challenge the results of VERIPOL This implies that the procedure must guarantee effective access to information on the operation of the tool and allow the affected party to exercise his or her right of contradiction, being able to present counter-evidence or request a review of the facts.

4.10. Neutrality and absence of bias in the use of VERIPOL

One aspect of legal relevance is the neutrality of the algorithms used by VERIPOL and the absence of discriminatory bias. As an AI tool, VERIPOL must be completely neutral and not introduce biases that unfairly affect certain social groups or categories of whistleblowers.

In this sense, it is the responsibility of law enforcement and judicial authorities to ensure that VERIPOL algorithms do not perpetuate pre-existing stereotypes or biases in the criminal justice system.[28] To prevent this type of situation, it is essential that the tool be subject to periodic audits and performance reviews to ensure its correct functioning and that the training process of the algorithms is carried out with representative data and without implicit biases. In addition, the judicial system must be in a position to correct any errors that may arise from unjustified bias in the use of VERIPOL.

4.11. Limitations on the use of VERIPOL: guarantee of non-decisional automation

The General Data Protection Regulation (GDPR), in Article 22, establishes a general prohibition on decisions based solely on automated data processing, including profiling, if such decisions have adverse legal effects or significantly affect the individual. This has direct implications on the use of VERIPOL, since its implementation cannot lead to automatic decisions that directly affect the complainant, such as the filing of the complaint or the initiation of criminal proceedings for false reporting.

To comply with this regulation, it is essential that VERIPOL be used only as an auxiliary tool[29] in the investigative process and that final judicial or police decisions be made by human beings.

---

[26] According to art. 120 of the Spanish Constitution. Judicial decisions must be motivated, exhaustive and congruent.

[27] Principle enshrined in art. 24 of the Spanish Constitution.

[28] For example, it could be the that complaints filed by certain vulnerable groups (such as immigrants or people with low economic resources) are disproportionately labeled as false due to linguistic patterns or social circumstances. This would be a violation of the principles of equality and non-discrimination (Art. 14 of the Spanish Constitution).

[29] The authorities should be empowered to review, interpret and, if necessary, rectify the results obtained from the tool, ensuring that the complainant has the opportunity to be heard and to present his or her allegations.

4.12. The right to explicability of algorithmic decisions

The right to explainability is an emerging concept in the field of the use of artificial intelligence in judicial and administrative processes. This right implies that persons affected by automated decisions, such as those generated by VERIPOL, should have access to a clear and understandable explanation of how the decision or recommendation was reached. In this sense, law enforcement authorities using VERIPOL must ensure that complainants can understand the criteria and patterns that the tool has used to label their complaint as possibly false. Lack of explainability could generate a risk of arbitrariness and opacity, which would undermine confidence in the justice system and could lead to the violation of fundamental rights. To avoid this, it is important that the authorities responsible for the use of VERIPOL implement mechanisms to audit and review the results and explain the algorithmic processes to affected citizens in an accessible and transparent manner.

5. Conclusion

VERIPOL is presented as an innovative tool that significantly improves efficiency in the detection of false reports, without compromising the legal guarantees of whistleblowers. Its use allows the optimization judicial and police resources, while protecting the fundamental rights of citizens and ensuring greater effectiveness in the prosecution of real crimes.

The tool, framed within the growing use of artificial intelligence in the field of justice, opens the way to a future where technology will decisively complement the work of the authorities, always in line with respect for fundamental rights.

In this paper we have focused on analyzing the main legal and technical aspects surrounding VERIPOL as they are multiple and complex, due to the innovative and sensitive nature of this type of tools in the criminal context.

From our technical point of view, we have concluded that the underlying architecture of VERIPOL, as well as the technological elements that comprise it, constitute a constantly evolving tool that is harmoniously aligned with contemporary advances in technology. As time progresses, the corpus of data feeding this system is expected to expand exponentially, which will allow the neural network to increase its learning and generalization capacity by incorporating an increasing number of test cases and analysis situations.

The progressive sophistication of processors, which will result in higher processing speed and a significant increase in the number of cores, will further enhance the efficiency of this tool.

This phenomenon will manifest itself in a substantial reduction of error margins, making VERIPOL an increasingly robust and reliable system.

In addition, the development of intelligent interfaces will open up a range of possibilities by integrating biometric and behavioral factors, such as assessment of an individual's nervousness, retinal analysis, heartbeat monitoring and neural communications. This multidimensional advancement in technology, similar to that seen in multiple contemporary systems, suggests that the combination of various intelligent

interfaces will not only enrich the capabilities of the system, but also significantly increase its accuracy and effectiveness in detecting anomalies and deceptive patterns.

In short, the synergy between the expansion of the database, the improvement of machine learning algorithms and the implementation of biometric interfaces converge to raise the quality and efficiency of VERIPOL, consolidating its position as an essential tool in the analysis and verification of allegations.

From our point of view, we consider that the use of artificial intelligence in justice must be strictly regulated and controlled to ensure that the established legal framework is respected, protecting the fundamental rights of citizens and avoiding possible arbitrariness or discrimination.

VERIPOL should be used exclusively as an auxiliary instrument or support tool, respecting the principles of proportionality, presumption of innocence, free assessment of evidence, and data protection. Its implementation and use must be accompanied by periodic audits for review and improvement, transparency in the implementation processes and absolute respect for the rights and procedural guarantees established by our legal system and governing criminal law.

References

Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. *O'Reilly Media*, Inc.

Cabrera, L., Fernández, I., & Gutiérrez, F. (2019). Natural language processing and automation in the detection of falsehoods in police reports: An evaluation of VERIPOL. *Forensic Linguistics Journal*, v. 15, n. 3, p. 56-67.

Camacho-Collados, J., Pilehvar, M. T., Collados, A., & Navigli, R. (2016). Supervised word sense disambiguation for multidomain content in public security. *Journal of Artificial Intelligence Research*, v. 57, p.449-473.

Delgado, D., Velasco, F., Pérez, J., & Gálvez, R. (2017). VERIPOL: Automatic analysis of fraudulent statements in the Spanish police. *IEEE Transactions on Computational Social Systems*, v. 4, n. 3, p. 132-140.

García, L., Pérez, M., Sánchez, A., & Gómez, P. (2018). Analysis of linguistic patterns in false reports using VERIPOL. *Journal of Police Studies*, v. 6, n. 2, p. 78-92.

Hernández, J., Ruiz, F., Martínez, S., & López, M. (2019). Artificial intelligence applications in policing: The case of VERIPOL. *AI and Law Enforcement Review*, v. 2, n. 1, p. 45-60.

López, M., Ortega, R., & Ruiz, C. (2018). Continuous Learning Models for Fraud Detection in Law Enforcement. *ACM Transactions on Knowledge Discovery from Data*, v. 12, n. 4, p. 27-35.

Martínez, A., Delgado, D., & Pérez, J. (2018). Machine Learning for Detecting False Reports in Law Enforcement: The Case of VERIPOL. I*nternational Journal of Artificial Intelligence and Law*, v. 24, n. 4, p. 269-282.

Méndez, J., Ortiz, A., & González, P. (2017). Automation in policing: The impact of VERIPOL on efficiency and accuracy in falsehood detection. *Law and Technology Quarterly*, v. 9. n. 2, p. 112-127.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.; Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, 2825-2830.

Quijano-Sánchez, L. Liberatore, F. Camacho-Collados, J, Camacho-Collados, M (2018). *Applying automatic text- based detection of deceptive language to police reports*: extracting behavioral patterns from a multi-step classification model to understand how we lie to the police. Knowledge-Based Systems, v. 149, p. 155-168.

Ramírez, P., Torres, J., & Gálvez, R. (2017). Evaluation of the VERIPOL system in the detection of false reports. *Procedia Computer Science*, v. 101, p. 134-141.

Rodríguez, S., Fernández, R., & Torres, V. (2018). Predictive analytics applied to whistleblower fraud detection: A VERIPOL-based approach. *Journal of Predictive Analytics in Law*, v. 11, n. 3, p. 91-104.

— V —

UPDATING NEURORIGHTS

— 16 —

The automation of the Judiciary: to what extent new technologies can put fundamental and social rights at risk. The centrality of the person.

FRANCISCO LUCIANO LIMA RODRIGUES
Universidade de Fortaleza
lucianolima@unifor.br

> "Technologies subject the concept of person to cross-oscillations of dilation and compression. Electronics comes to the conclusion that "we are our information"; Genetics, on the other hand, insists that "we are our genes". The technological emphasis will have to be mitigated to prevent biology from becoming a biography and virtuality from dragging people down new paths of abstraction. **The person will always be something more than the set of physical and virtual data that composes him**." **Stefano Rodotà**, **in** The Right to Have Rights. p.159-160. Editora Trotta, S.A, Madrid, 2014. (free translation)

Introduction

The digital transformation of the Brazilian Judiciary has been an irreversible and progressive process, driven by the advancement of new technologies and the search for greater efficiency and transparency in the provision of jurisdiction. The use of Artificial Intelligence (AI), electronic systems, and digital platforms for procedural processing represent a revolution in the way justice develops its activities, bringing numerous benefits, such as greater speed in the analysis of processes, reduction of bureaucracy, and improvement of judicial management. However, this technological revolution raises fundamental questions about its impacts on fundamental rights and

the centrality of the person, a principle guaranteed by the Federal Constitution and which should be the structuring axis of any innovation applied to the justice system.

The centrality of the person, understood as the primacy of the person in any other situation, especially in the market, in relation to their rights, dignity and individual guarantees related to automated processes, should be the essential criterion for evaluating the incorporation of new technologies into the Judiciary. The risk of dehumanization of justice, the loss of observation of the uniqueness of the person in concrete cases, and the difficulty of challenging automated decisions are just some of the concerns that emerge with the use of AI in the judicial sphere. When algorithms start to perform functions traditionally assigned to judges, lawyers, and civil servants, there is a need for careful regulation to ensure that such tools are used as a support, and not as a substitute for human activity.

Another critical challenge addressed in this research is the digital divide and its repercussions on access to justice. The computerization of the Judiciary, while expanding accessibility for some, imposes barriers for those who do not have the technological resources or knowledge necessary to use electronic systems. Vulnerable populations, such as low-income people, the elderly, individuals with disabilities, indigenous communities, and residents of rural areas, face difficulties in interacting with digital platforms, which can result in the restriction or even denial of their right to access justice.

In view of this panorama, this article proposes a critical analysis of the automation of the Judiciary, emphasizing the need for the use of technology to respect the centrality of the person and not be limited only to effecting institutional efficiency. To this end, the article examines the impacts of AI on fundamental rights, the challenges of the digital divide, and the risks of standardizing judgments by algorithmic systems. It also discusses the measures that must be adopted to ensure that technological advancement in the Judiciary respects human dignity, democratic principles and social rights, ensuring that innovation does not result in setbacks in the rights conquered.

1. Digital transformation within the Judiciary

The use of electronic instruments by the Judiciary in Brazil had as its starting point Law 8900, of May 26, 1999 (Brazil, 1999), which allowed, in its article 1, the parties to use a data and image transmission system of *facsimile* type or other similar, for the practice of procedural acts that depend on a written petition.

This Law became known as the "Fax Law", changed the concept of paper proceedings, allowing both parties and magistrates to use an electronic instrument to forward documents to the proceedings. However, the need to attach the physical document, within five days of sending it electronically, indicated a certain distrust of the use of the electronic novelty.

Another important moment in the trajectory of the Brazilian Judiciary towards the digital process was the enactment of Law 14.419, of December 19, 2006 (Brazil, 2006), which provided for the computerization of the judicial process.

This Law allowed the use of electronic proceedings in the civil, criminal, and labor areas, as well as in the Special Courts, by the use of a digital certificate for digi-

tal signature. Among other changes, the Electronic Justice Gazette was implemented, available on the World Wide Web, to publish judicial and administrative acts; the method of counting deadlines when using the electronic process was established; allowed the bodies of the Judiciary to develop electronic systems to be used in judicial proceedings through digital records; indicated the need to use open source, accessed through the World Wide Web, prioritizing the standardization of the systems.

With the implementation of the electronic judicial process, without a doubt, many advantages have been brought to the judicial services. However, with the novelty, there were also problems, both of a technical nature and others generated by the Brazilian socioeconomic reality that exhibits an abyssal social inequality.

The National Council of Justice, the administrative and financial control body of the Judiciary, created by Constitutional Amendment 45, of December 30, 2004 (Brazil, 2004), hired a consulting service, in partnership with the Institute of Technology and Society (ITS), which was prepared by the School of International and Public Affairs (SIPA), at Columbia University, called "The Future of AI in the Brazilian Judicial System" in which it set three objectives: i) to create a panorama (capable of being implemented later) to map and categorize the various artificial intelligence (AI) tools already developed by the Brazilian Judiciary, including a comparative study and an integration and standardization model; ii) design a collaborative governance structure that allows the judiciary to achieve greater cooperation and collaboration; and that it works in a manner compatible with the Electronic Judicial Process (PJE) and its tools; iii) conduct an analysis of the principles, incentives and internal regulations that govern the operation of the PJE Innovation Laboratory (Electronic Judicial Process), including a proposal for improvement and expansion of current management models, in accordance with international best practices.

Also in this document, it is observed that the consultancy carried out establishes an Agenda with the principles that must be observed by the use of Artificial Intelligence (AI) by the bodies of the Judiciary, among them, the role of human supervision and human rights.

With regard to human rights and security, the document produced by Columbia University and the Institute for Technology and Society (ITS) points to concern when it states; "[…] Another relevant point is the concepts of personal data and privacy protection. Google's *white paper* on AI discussed the importance of comprehensive data protection legislation in order to mitigate the human rights risks posed by AI. However, they also warn that additional measures are necessary to protect personal data. Brazil has the General Data Protection Law (LGPD), which deals with the use of personal data in the public and private sectors, as well as stipulates under what circumstances (such as consent) an entity is authorized to use it. The CNJ should also rigorously examine the types of personal data being used by AI tools and take additional measures to ensure data protection."

The Justice 4.0 Project, promoted by the National Council of Justice in partnership with the United Nations Development Program 32345'678 (UNDP) aims, in summary, to provide the entire Brazilian Judiciary with the digital process in order to maximize productivity, reduce bureaucracy and integrate information.

The use of the digital process by the Brazilian Judiciary reaches almost the entire procedural collection that is processed in the Brazilian Courts. However, the

modernity brought by technology has not been able to avoid the digital divide that affects a considerable portion of the Brazilian population, keeping the poorest away from access to justice, with serious offense to human rights.

2. Artificial intelligence and fundamental rights

For the development of this work, it is essential to carry out an analysis of the definition of Artificial Intelligence, with a brief report on its origin and development.

Likewise, the definition of fundamental rights is indispensable to achieve the intended objective, especially with an approach to those related to the digital context.

2.1. Concept of Artificial Intelligence

In order to define the concept of Artificial Intelligence (AI), it is possible to start from the conceptualization attributed by bills that aim to regulate the use of AI, as well as from international regulations and also from doctrine.

In Brazil, Bill no.2338/2023 defines an AI system, in the following terms:

> Art. 4 For the purposes of this Law, the following definitions are adopted:
>
> I – artificial intelligence system: computational system, with different degrees of autonomy, designed to infer how to achieve a given set of objectives, using approaches based on machine learning and/or logic and knowledge representation, through input data from machines or humans, with the objective of producing predictions, recommendations or decisions that may influence the virtual or real environment (Brazil, 2023);

The definition contained in PL2338/2023 was subject to consideration by the Committee of Jurists[1] established to analyze the bills that aim to regulate the use of Artificial Intelligence in Brazil and, among its members and other participants, there was no convergence regarding the definition of AI.

The aforementioned bill was considered in December 2024 by the Federal Senate, which approved a substitute for the original bill and forwarded it to the Chamber of Deputies for consideration, as part of the legislative process.

Among the many modifications to the original project, the following stand out, due to its importance, the maintenance of social network algorithms and the provision that provides for the protection of the rights of creators of content and artistic works out of the list of systems considered to be of high risk.

Regarding the conceptualization of Artificial Intelligence, say Barrea, Salmoria and Rodrigues (2023, p. 333)

> To form the concept of AI, it is necessary to relate it to other areas of knowledge, as it reaches all segments of society. AI technology is used to read exams, grant credit, geolocation, streaming services, among others.
>
> In fact, Artificial Intelligence is composed of models, created by programmers, which select the information to be provided to the system and that will be used to predict solutions and results.
>
> From data and information, human activity treats this material, so that facing the (im)possibility of neutrality of those who produced them constitutes a concern with the results (outputs) that, depending on the programmers who insert the information, are covered with cognitive biases.

[1] Committee of Jurists established by the Act of the President of the Senate No. 4, of 2022, aimed at subsidizing the preparation of a draft substitute to instruct the consideration of Bills No. 5,051, of 2019, 21, of 2020, and 872, of 2021, which aim to establish principles, rules, guidelines, and foundations to regulate the development and application of artificial intelligence in Brazil.

The European Union recently issued the first regulation on Artificial Intelligence (AI), known as the AI Act, whose discussion began in 2021, with the European Commission's proposal that aimed to establish a harmonized legal framework for the development and use of AI in the European Union.

The regulation was approved in May 2024, with entry into force on August 1 of the same year, and its gradual application was established over the six to thirty-six months following the entry into force of the *IA Act*.

Among the fundamental points of AI regulation, the following stand out: i) the definition of an AI system, aligning with the approaches of the Organization for Economic Cooperation and Development (OECD) to differentiate AI from simple software; ii) the adoption of a risk classification, pointing to AI systems as unacceptable risk, high risk, limited risk and minimal risk; iii) the creation of a governance and inspection system, with the purpose of supervising the implementation of the rules, especially for general-purpose AI; iv) the prohibition of practices such as cognitive behavioral manipulation, emotion recognition in work environments and social *scoring*; v) the observation, by implementers of high-risk AI systems, of the impacts on fundamental rights before putting these systems into use.

This regulation of AI by the European Union, since the first negotiations, has served as a parameter for many countries, including Brazil, in the discussion and regulation of the matter, especially with regard to the conceptualization of AI and the preservation of fundamental rights, democracy and the rule of law.

The European Union has issued Regulation (EU) 2024/1689 of the European Parliament and of the Council, dated 13 June 2024, with harmonised rules on artificial intelligence, which in its considerations states:

> The objective of this Regulation is to improve the functioning of the internal market by providing for a uniform legal framework, in particular for the development, placing on the market, putting into service and use of artificial intelligence systems (AI systems) in the Union, in line with Union values, in order to promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union ('the Charter'), including democracy, the rule of law and protection of the environment, protection against the harmful effects of AI-systems in the Union, and supporting innovation. This Regulation ensures the free cross-border movement of AI-based products and services, thereby preventing Member States from imposing restrictions on the development, marketing and use of AI-based systems, unless explicitly authorised by this Regulation.

It can be seen that the European Union's rules regarding Artificial Intelligence are concerned with preserving the centrality of the person, trust, a high level of protection for health, safety and, with special emphasis on fundamental rights, as well as democracy, the environment and the rule of law.

The concerns raised in the European Union document are also present in Bill 2338/2023, approved by the Federal Senate and now under analysis by the Chamber of Deputies, thus observing the commitment of the Brazilian parliament to the centrality of the person and the protection of fundamental rights, when regulating Artificial Intelligence in Brazil.

In Regulation (EU) 2024/1689 of the European Parliament and of the Council, article 3, item I, defines an Artificial Intelligence System, in the following terms:

> 'AI-system' means a machine-based system designed to operate with varying levels of autonomy, and which may exhibit adaptability after deployment and which, for explicit or implicit purposes, and on the

> basis of the input data it receives, infers how to generate outputs, such as predictions, content, recommendations or decisions that may influence physical or virtual environments;

It is noted that both in the definition contained in Regulation (EU) 2024/1689 of the European Parliament and of the Council, and in Bill 2338/2023, machine learning, including supervised, unsupervised and reinforcement learning; systems based on logic and knowledge; statistical methods, in addition to the definition of related terms such as input data, biometric identification, and biometric verification, among others, to cover detailed aspects of the use and application of AI.

Thus, any definition of an Artificial Intelligence System must contain, in addition to the requirements common to its description, mechanisms that guarantee a safe, ethical development that respects fundamental rights.

2.2. Fundamental rights

Fundamental rights are inherent to the human condition itself, linked to human rights. Throughout history, several moments have been important for the construction of what is called fundamental rights, with emphasis on the Magna Carta of 1215 where, for the first time, limits were placed on the absolute power of the monarch and rights were established for subjects, among them, what is now called due process, the right to a fair trial and the prohibition of confiscation.

Next, the Virginia Declaration of Rights (1776) and the United States Declaration of Independence (1776) can be indicated as instruments for the constitution of fundamental rights, which affirmed inalienable rights, such as life and liberty.

The Bourgeois Revolution of 1789 marked the rupture of a relationship of vassalage, with limitation of rights, to give rise to a new legal, economic and social order, marked by the tripod – equality, liberty and fraternity, added to the right to private property, giving rise to the Declaration of the Rights of Man and of the Citizen, whose content guaranteed, in addition to bourgeois ideas, freedom of expression, equality before the law and the presumption of innocence.

Fundamental rights have also been guaranteed in several modern constitutions, among them, the Constitution of the United States (1787) guarantees fundamental rights such as freedom of speech, religion, press, assembly and petition, in addition to protecting against arbitrary searches and seizures, guaranteeing due process and the right to a fair trial, the French Constitution (1791)

Declaration of the Rights of Man and of the Citizen (1789), incorporated into the French Constitution of 1791, establishing fundamental rights such as liberty, equality, property, security and resistance to oppression and the Mexican Constitution (1917) resulting from the Mexican Revolution that came to incorporate social rights, including rights to work, education, health and agrarian property, as well as civil and political rights.

The Weimar Constitution marked the combination of classical civil and political rights with social and economic rights with the aim of establishing a more just and balanced society.

Among the main fundamental rights guaranteed by the Weimar Constitution, the following stand out: the right to equality before the law, without discrimination of race, origin, language, class or religion; freedom of the press and opinion; freedom

of association and peaceful assembly without prior permission; the right to privacy; universal suffrage; It guaranteed due process of law and also the protection of private property, with the condition that it must serve the common good, and that expropriations can only occur with fair compensation.

The atrocities of the Second World War made the most degrading side of the human being known, with the most demeaning demonstrations of evil, but, on the other hand, there was a reaction in the sense of placing dignity as an inherent condition of the human being.

In the post-war period, the United Nations was established and then the Universal Declaration of Human Rights (UDHR) was promulgated, which is a milestone in the universal protection of human rights.

The United Nations, regarding the effects of the Universal Declaration of Human Rights, states:

> The UDHR, together with the International Covenant on Civil and Political Rights and its two Optional Protocols (on the complaint procedure and on the death penalty) and with the International Covenant on Economic, Social and Cultural Rights and its Optional Protocol, form the so-called International Bill of Human Rights.
>
> A number of international human rights treaties and other instruments adopted since 1945 have expanded the body of international human rights law.
>
> They include the Convention on the Prevention and Punishment of the Crime of Genocide (1948), the International Convention on the Elimination of All Forms of Racial Discrimination (1965), the Convention on the Elimination of All Forms of Discrimination against Women (1979), the Convention on the Rights of the Child (1989) and the Convention on the Rights of Persons with Disabilities (2006), among others.

The emergence of new technologies, from the second half of the last century, including computers and artificial intelligence, have put fundamental human rights at risk, since technological innovations have shown signs that they could reduce or even annul rights to freedom of expression, privacy, driving the need to protect fundamental rights in the digital age.

Thus, together with the technological advances related to Artificial Intelligence, it is essential to adopt measures related to the protection of fundamental rights, especially with regard to the dignity of the person, so that new technologies can be enjoyed, but without allowing setbacks related to conquered rights, especially the right to privacy, freedom of expression, non-discrimination and above all, the right to information on how their data is used by AI, always with the centrality of the person prevailing.

3. Impact of AI on Fundamental Rights

The emergence of new technologies, including Artificial Intelligence, has frequently had an impact on fundamental rights, either due to the opacity of algorithms, or due to moderation control exercised without human participation, as well as due to the influence of the market in a society where its strength tends to prevail over the centrality of the person, causing injuries to the right to privacy. the protection of sensitive personal data, freedom of expression and information, the right to non-discrimination and equality, among others.

3.1. Privacy and Data Protection

The Brazilian constitutional order, whose foundations are set forth in article 1 of the Federal Constitution, ensures the dignity of the human person, as well as ex-

plicit in the list of fundamental rights, contained in article 5, to be ensured, the right to privacy and data protection.

The General Data Protection Law (LGPD) establishes:

> Article 1 This Law provides for the processing of personal data, including in digital media, by an individual or by a legal entity governed by public or private law, with the objective of protecting the fundamental rights of freedom and privacy and the free development of the personality of the natural person.
>
> Art. 2 The discipline of personal data protection is based on: I - respect for privacy; II - informational self-determination; III – freedom of expression, information, communication and opinion; IV – the inviolability of intimacy, honor and image; V – economic and technological development and innovation; VI – free enterprise, free competition and consumer protection; and VII – human rights, the free development of personality, dignity and the exercise of citizenship by natural persons.
>
> Article 5 For the purposes of this Law, the following shall be considered:
>
> I - personal data: information related to an identified or identifiable natural person;
>
> II – sensitive personal data: personal data on racial or ethnic origin, religious conviction, political opinion, membership in a union or organization of a religious, philosophical or political nature, data related to health or sex life, genetic or biometric data, when linked to a natural person; [...]

Bill 2338, of 2023, which provides for the use of Artificial Intelligence, points out in its article 2, the foundations for the implementation and use of AI systems, including the centrality of the person and privacy, data protection and informational self-determination.

Although the right to privacy and data protection are enshrined in the Federal Constitution as fundamental rights and protected by the General Data Protection Law (LGPD), the emergence of new technologies, such as Artificial Intelligence (AI), has produced a series of violations of these rights, especially by the way this data is treated, importing privacy exposures and data misuse.

In this regard, Linera (2022, p. 37), when dealing with the issue within the framework of the European Union, states:

> It seems that among the fundamental rights most affected by the use of AI systems will be those that guarantee the private dimension of individuals: privacy, one's own image, the protection of personal data and the secrecy of communications, recognised in paragraphs 1 and 4 of Article 18 EC, in Article 8 ECHR and Articles 7 and 8 of the Convention on the Functioning of the European Union. and one of the factors that make AI such a powerful tool is its capacity to process a huge amount of data and to detect the connections that may exist between them, as well as its potential to, as recalled in the European Union White Paper on AI (p.15), *"track and deanonymize"* data relating to individuals, and thus generate new risks around the protection of personal data in relation to data sets that do not contain personal data and that is that consent to the transfer of two or three apparently irrelevant data can generate, when inserted within a much larger set of data, "baby data" unpredictable for the person concerned, which has a very important effect on requirements such as consent, expression of personal autonomy and transparency.

The guarantee of privacy and data protection, through the use of Artificial Intelligence, in order to meet the constitutional precepts that place them as fundamental rights, requires the taking of measures, on the part of those who, not being holders of these rights, do not use them without due prior consent, including the impossibility of changing or canceling them.

The advantages of using AI to process personal data cannot be ignored, and the assertion that the generation of data produced by it can be faster and more efficient than that carried out manually by humans is true, as well as that the use of this data by the public administration can favor the reduction of errors and accelerate results.

The issue is not restricted to whether or not to use AI, but to place as a limit for its use, respect for fundamental rights, preventing the use of personal data, with the

exposure of the privacy of its holder, from bringing harm to the person or that it is used to enrich large companies that operate in the field of new technologies.

Also among the risks of using AI in disagreement with fundamental rights, it can be pointed out that the autonomy of data processing, especially through the use of algorithms, undermines the person's right to informational self-determination, through improper use or beyond what is consented.

The challenge that arises in relation to the use of AI and the fundamental right to privacy and data protection is to regulate the use of technology in order to preserve and protect such rights, without impeding the development of new technologies, that is, the balance between the advancement of technology and the centrality of the person.

3.2. Freedom of Expression and Information

Freedom of expression is one of the most important achievements of the French Revolution, when the relationship of vassalage that marked the Middle Ages ended, with the consequent issuance of the Declaration of the Rights of Man and of the Citizen, of 1789, ensuring equal treatment between men, freedom of belief and freedom of expression and information.

The right to freedom of expression and information has been one of the great challenges presented by the use of Artificial Intelligence, considering the scenario in which new technologies and social networks have expanded the space for people's manifestation, with greater ease, reach and speed.

Traditional communications carried out by radio, television and newspapers, which had already suffered a decrease in their reach with the introduction of the World Wide Web in the last century, were further reduced with the emergence of AI, whose greatest advantage over traditional models would be in the ease of manifestation, the speed of its transmission and the processing of data.

Among the objectives of the Republic, listed in article 3 of the Brazilian Federal Constitution, there is political pluralism, the basis of the Democratic State, translated by the guarantee of the diversity of political positions, as well as by the possibility of defending them freely, without limit and without the imposition of controls related to content.

The use of AI has often resulted in the exposure of sensitive data, those related to it described in article 5, I, of the General Data Protection Law (LGPD),[2] especially when the selection of data and its use happens through the use of an algorithm.

Freedom of expression in its relationship with Artificial Intelligence can offend fundamental rights to the extent that the search for information on social media can be influenced by the algorithm that will only lead to certain information that serves the interests of economic or political groups, there would therefore be a violation of the fundamental right of expression, To the extent that receiving information cut down by the algorithm, in order to meet a certain interest, freedom of expression would be compromised, as a consequence.

[2] Article 5 For the purposes of this Law, the following shall be considered: […] II – sensitive personal data: personal data on racial or ethnic origin, religious conviction, political opinion, membership in a union or organization of a religious, philosophical or political nature, data related to health or sex life, genetic or biometric data, when linked to a natural person; […] (Brasil, 2018).

By the way, Arroyo (2022), states:

> Large search engines are the instruments for searching, receiving or imparting information; however, their content is indexed or classified by algorithms that can lead to information fragmentation, favoring only certain platforms and types or conglomerates of media. In this sense, the algorithm may also be biased towards certain types of content or content providers, with the danger of affecting values related to pluralism and information tolerance of various media.

The use of artificial intelligence and algorithms in the production of information and, consequently, in the exercise of the right of expression, are liable to be subjected to moderation mechanisms, as a way to avoid violations of fundamental rights. It happens that, many times, there is no knowledge of how the moderation process is developed and applied, leaving doubts about human participation and the use of algorithms. In this regard, Silva (2022) says that little is known about moderation practices on platforms and social media and, often, when it is known, it is through press leaks. It also argues that this hard work of moderation of billions of units of content is distributed geographically and racially in the world, with rules established by major technology professionals, aiming at profitability and the escape from legal implications, carried out, almost always, by precarious professionals from countries in the global South.

The socioeconomic inequality that exists in the global South, so patent and clear in the areas of health, education and subsistence conditions with a minimum of dignity, also appears in identical proportions in the influence that large economic groups exert on the use of artificial intelligence and algorithms to the detriment of the fundamental rights to freedom of expression and information. reaffirming the tendency that makes the market prevail over the uniqueness of the person.

3.3. Non-discrimination and equality

Addressing the relationship between Artificial Intelligence and the fundamental right to non-discrimination and equality it is important to bring to the discussion structural racism, whose verification is presented by the perpetuation of racism through social, political and economic systems and institutions, materialized by systemic disparities in areas such as education, health, housing, employment and criminal justice, where particular racial groups face significant and persistent barriers.

Regarding the origins of structural racism in Brazil, one of the most expressive factors can be indicated as the formation of the Brazilian state, founded on a patriarchal, prejudiced and slave-owning society, marked by an extensive period of discrimination that gave rise to a racial inequality that has been reproduced for generations and that can be seen in the absence of investment in health, housing, schools, and public sanitation in areas occupied mostly by black people, not to mention the lack of access to justice that make clear the disparities in the justice system in the disproportionate rates of incarceration of black people and the underrepresentation of minorities in positions of power and influence as patent examples of structural racism.

Regarding *online racism* in Brazil, Silva (2022) points to research developed by Luiz Trindade (2018) that brings especially relevant data on the behavior of aggressors and characteristics of the targets: 81% of the victims of racism on Facebook in Brazil are middle-class black women, with higher education; 76.2% of the aggressors had no previous relationship with the victim; 66% of the aggressors are young;

and often the articulations to Racist slurs are reactions to positive events exposed on social media by black women.

Violations of the fundamental rights of non-discrimination and equality can also be detected in the so-called algorithmic discrimination that translates into behavior where artificial intelligence systems perpetuate and amplify pre-existing prejudices and social inequalities.

Considering that algorithms, as a rule, are conditioned by the human being, the biases that can produce discrimination arise from biased historical data or inadequate selection of training data, resulting in systematic negative impacts on minority groups in areas such as employment, credit, criminal justice and access to services.

In this regard, Noble (2018, p. 1) states that algorithms reflect the priorities, values, and judgments of those who have the power to encode them, often reinforcing and perpetuating social inequalities.

With the same concern, Arroyo (2022) warns about the risk of decision-making algorithms, based on previous human decisions, marked by the same biases, which already hinder daily decision-making and which will probably be replicated and multiplied in algorithmic systems. As an example, the aforementioned author points out that a biased algorithm that systematically discriminates against a group in society (e.g., based on their age, sexual orientation, race, gender, or socioeconomic position) can raise considerable concerns, not only socially, but also by the use of voter preference and behavior algorithms that can suffer an equal or more powerful impact than traditional media.

In this way, the use of artificial intelligence, as well as algorithms, when not subordinated to the limits of respect for fundamental rights, the centrality of the person and the primacy of equality, can establish discrimination of different orders and become even more worrying when, due to their biases, they perpetuate structural racism.

4. The centrality of the person guaranteed by the Federal Constitution

The analysis of the constitutionalization of law modifies the old nineteenth-century understanding of civil law, with a patrimonialist and individualist bias, by which property and contract were at the top of protection, the thing, therefore, was placed to the detriment of the person. It is the conception, already outdated, that civil law takes care of and resolves issues involving individual interests and constitutional law, questions related to the organization of the State would remain. It would be dualism – public law *versus* private law. The State taking care of the collective interest, the person of its interests. Each one in different fields, in circles with almost or no intersection.

The liberal state, with the evidence in the individual, made clear the presence of a subject of law - the abstract subject, marked by the privilege of bearing the condition of owner and, therefore, that of contracting, being called subject-owner. This liberal idea intended to fulfill the determination achieved by the ideals of the French Revolution, especially that of equality.

However, the equality defended by liberals did not recognize the singularity and plurality of the person which, in the words of Hanna Arendt (2009), would be the paradoxical plurality of singular beings.

The abstract subject, however, molded in excessive formality, was not capable of understanding reality, of admitting and protecting the differences between people, his perspective was closed within what was codified, initially in the French Civil Code of 1804, later spreading to other foreign legislations, including the Brazilian Civil Code of 1916, where the pretension of completeness and self-sufficiency, It ended up essentially regulating rules that served the owner subject, excluding women, minors, people with disabilities, blacks and other vulnerable groups.

The proposal of the liberal state, with the subject-owner, contained the idea of the "world of certainties", everything was provided for in the codified legislation, there was the materialization of the principle of equality, in this case, formal equality.

In this scenario, where the abstract subject prevails, jurists began to realize that the idea of abstraction was no longer presented, as Rodotá (2014) states, as an instrument capable of understanding the realized, but, on the contrary, it was an impediment, an obstacle, imposing the annulment of the subject by the loss of the face, as in the figures portrayed in the works of Francis Bacon, where the body is present, but the vision of the face is blurred, smoked, without particularities that differentiate it from the others.

There was the prevalence of the thing, of the abstract subject, without a face, over the person. This situation could be observed in the treatment given by the Brazilian Civil Code to women, whose expression of will was at a lower level than that of the husband who prevailed; excessive concern with the assets of the wards to the detriment of their well-being; the exercise of the right to property in an absolute way, disregarding its effects on the environment and people's health; the denial of full capacity to people with disabilities, in addition to the little concern for the vulnerable in general.

Legal relations in the liberal state were founded on economic freedom, on the "invisible hand" of the market, as stated by Vieira de Andrade (2001), which, in the prevailing view at the time, would lead to the best of all worlds, but, in reality, the intention was to avoid interference by the State that could reduce the performance of the "invisible hand", putting at risk the values of autonomy of the private sphere. the freedom and property of individuals.

The man of the Liberal State was reduced to the condition of a subject of law and, as such, had no feelings, he was moved only by the law of the market. A "metalist automaton", his motivations centered, as Aronne (2013) states, on the theory of fair exchange.

However, the effects of the First World War did not take long to influence the flexibility of the concept of completeness and self-sufficiency of civil law and, thus, to put in check the figure of the subject of law. In Germany, says Wieacker (2010), it proved necessary to impose restrictions on freedom of contract and the freedom to use property, giving rise to a global economic policy that began to influence private law, with the emergence of a constitution that, in the event of the failure of free competition or the threat to social solidarity, gave rise to an intervention by the public authorities with a directing function by the public administration of the economy.

In Brazil, following the 1930s, there was also a breakdown in the feeling of completeness of the Civil Code, based on greater State intervention in the economy, with the imposition of restrictions on private autonomy and the presence of the phenomenon called contractual dirigisme. At this moment, it is observed that the idea of absolute property so characteristic of the liberal state, power and strength of the subject-owner, suffers fissures with the distribution of the property of the soil and subsoil, imposed by the Federal Constitution of 1934, as well as the edition of Decree-Law 25, of 1937, still in force, which regulates the preservation of assets of cultural interest through the institute of Listing, a limitation on the right to property.

Such phenomena favored a relocation of constitutional law, leaving a position of mere repository of political expectations, far from the daily life of the common man, whose centrality of his legal relations was located in property and contract, to place at the center of the legal system the constitution with all its load of principles to serve as a parameter for the interpretation of all the other norms existing in the legal system.

The dialogue between civil law and the constitution and, therefore, a displacement of the subject of law towards the person, based on the principles and objectives of the republic that should guide the interpretation of the norms, favored the rupture of the prevalence of the thing over the person and, in an efficient way, pointed out to the interpreter the important element to be preserved, in the case of family relations, the aspect of affectivity as opposed to patrimony, in contracts the existence of weak contractors, in private property the fulfillment of the social function, in the use of technology the preservation of their intimacy.

As already said, man within his plurality is unique in the sense that he must be the end and not the means, leading to the understanding that he cannot be used as an instrument for the personal projects of another, but that the State is the one that must serve the person and not the other way around, reaffirming the primacy of the importance of dignity as an intrinsic value of each person.

The inclusion of the dignity of the human person in international texts and constitutions, since the Second World War, has consolidated understandings about its content, even recognizing multiculturalism, the diversity of many ethnicities and societies, but extending its scope to respect for the physical, psychic and bodily integrity of the person, admitting that there are basic needs, without which, the dignity of the human person would be violated, among them, the guarantee of housing, basic health, education and work conditions.

In addition to meeting the basic needs of the person, it is necessary for the fulfillment of the dignity of the human person, the adoption of equal treatment that prevents any type of discrimination, whether economic, ethnic, cultural, sexual option, religious or political, whose fulfillment favors equal political participation. Equality is a prerequisite for meeting the minimum requirements of the primacy of the dignity of the human person, and it is inadmissible for a dignified existence of the person to treat him or her as an object.

In the Brazilian legal system, as of the current Constitution, the dignity of the human person is the foundation of the system, serving as the basis, parameter and limits of its whole, boasting a superior fundamentality in relation to the other principles, without discrediting the absence of hierarchy between the principles and their superiority over the rules.

In an unequal society such as the Brazilian one, conservative and patrimonialist, with an undisguised difficulty in recognizing social differences and combating them, the principle of the dignity of the human person is an effective instrument to prevent the objectification of man, and it is essential that the implementation of new technologies, whether by the Judiciary or other State institutions, does not lose sight of the fact that the centrality of the person is an inalienable constitutional imposition. under any justification.

5. The digital divide and access to justice

Access to justice is not restricted only to the possibility of filing lawsuits, as has been imagined for a long time. On the contrary, the current concept of access to justice has become broad, including not only the possibility of having resolved the pendency through a judicial decision, but, above all, because it has within its reach the Public Defender's Office for those who do not have resources, as well as free costs and emoluments, the resolution of conflicts by alternative forms and, more recently by the possibility of using the digital process.

The option of using the digital process by the Judiciary had as its scope greater accessibility to judicial services with speed and security. However, along with the benefits that the digital process has brought, there is an important aspect that deserves to be highlighted: the digital divide.

As the process became digital, a legion of people emerged who did not have access to the new model, either due to economic limitation or technical inability to handle the new tools.

At this moment, there is the presence of the Brazilian reality, arising from the formation of the Brazilian State, whose mark was patrimonialism and conservatism, with the exclusion of vulnerable groups that, in some way, did not serve the market and, therefore, had no interest or protection.

Brazilian socioeconomic inequality once again justifies the digital divide, making clear the affected groups, in this case, low-income people, the elderly, women, people with disabilities, indigenous populations, immigrants, residents of rural areas, all those who face barriers that put their fundamental rights at risk.

The high rate of illiteracy, unemployment, low education, and the difficulty of dealing with digital technologies, such as the internet and electronic devices, translate the digital divide that can jeopardize participation and involvement in spheres of interest, whether linked to public administration or access to the Judiciary.

The lack of minimum computer knowledge results in difficulties in accessing information and services made available by the Courts, in addition to the language used by the procedural systems.

The absence of public policies that make the contact of those who are unable to understand or handle the equipment of the digital judicial process more accessible, requires the Judiciary to facilitate access to justice by adopting language that allows the understanding of judicial acts. An example of facilitating the understanding of judicial acts performed by digital means is the use of *visual law* that clearly explains, using figures, videos or other mechanisms for which the procedural act for which it was requested is intended.

The elaborate language historically adopted by the Judiciary has always been a barrier to understanding for those with low education or reduced cognition, an aspect that was aggravated with the adoption of the digital process. There would now be two problems that would keep citizens away from access to justice – inaccessible language and the use of unknown digital tools.

It is essential that the Judiciary adopt measures that can guarantee the digital inclusion of vulnerable groups through a structure that allows access to judicial services for all who do not have the economic conditions to acquire electronic equipment, as well as for those who have limitations to use the digital process, offering access through traditional channels.

The digital divide has negative impacts on access to justice, as the impossibility of using digital technologies imposes a limitation of rights on vulnerable groups.

Among the negative aspects that may arise from the right of access to justice due to the digital divide, the following can be pointed out:

(i) The digitalization of processes: with regard to the security and ease of handling the records in any location, as long as they have access to the World Wide Web, it has brought limitations to vulnerable groups who, due to the absence of specific computer knowledge, are restricted in accessing the records and, thus, prevented from, for example, complying with procedural acts.

Not that the digitization of the records is a problem, quite the contrary, it facilitates access to judicial services. However, this advantage has only been enjoyed by a group of people who have knowledge of information technology and who can purchase equipment and hire data plans.

Those who are not in this group, the digitally excluded, are at the mercy of the electronic procedural systems adopted by the Courts or, when they do, they are precarious and do not lend themselves efficiently to alternative access for those users who, for some reason, cannot use electronic means.

(ii) The difficulty in communicating with the Courts has been another negative impact of the adoption of the digital process by the Courts, since communications from then on occur only by digital means, with the use of e-mails, communication platforms such as Whatsapp or Telegram, disregarding the social reality in which a considerable part of the population does not have the economic or understanding conditions to acquire or use digital tools.

Once again, it is observed that the adoption of digital models in judicial proceedings does not take into account vulnerable groups, and such behavior may result in harm to the excluded who, not of their own volition, may miss judicial deadlines and, as a consequence, have been denied the guarantee of due process of law for not participating in judicial proceedings.

6. Commitments that can be adopted by the Judiciary to combat the digital divide

The Judiciary is not a mere spectator in this scenario of adoption of digital processes, whether by civil society or by the public authorities, much less an institution indifferent to the consequences of the adoption of new technologies, but, above all,

committed to the centrality of the person, the preservation of his dignity, as one of the objectives of the Republic.

As the holder of part of the State's power, it is up to the Judiciary to ensure the effectiveness of the guarantees of ample defense, adversarial proceedings, whether in the judicial, administrative or private relations spheres, through digital inclusion, in order to ensure equitable access to technologies, favoring citizens, especially those who are vulnerable, the acquisition of skills to use digital technologies.

Also within the responsibility of the Judiciary to ensure equal access to the new technologies used by the justice system, it is necessary to provide face-to-face assistance services for those who do not have access to or are not familiar with digital technologies. To this end, the Judiciary must effectively maintain the qualification and training of civil servants to support vulnerable people in the use of new technologies and, thus, mitigate the digital divide.

The updating of data related to the advances and possible setbacks in digital inclusion related to judicial processes, requires the Judiciary to constantly carry out research and data collection and, in this way, subsidize decisions in order to maintain the growing adoption of measures that combat the digital divide.

The fight against digital exclusion requires the adoption of measures that ensure a more egalitarian and inclusive society, with the understanding that the subject cannot do without a holistic approach that considers the various reasons for its occurrence, with emphasis on poverty and discrimination in all its dimensions.

Therefore, the fight against the digital divide related to vulnerable groups requires, among other measures, the strengthening of the defense of human rights through public policies that guarantee respect for the dignity of the person, considering their uniqueness; the fulfillment of the objectives of the Republic indicated in article 3 of the Constitution of the Federative Republic of Brazil,[3] especially those that deal with the fight against poverty and regional inequality, in addition to the recognition of political plurality; Without forgetting access to education in all its forms, both formal and regular education offered by educational institutions, as well as other forms of education by offering access to inclusive schools, art and learning to be made available to all age groups.

In addition to all these measures, the fight against digital exclusion does not dispense with the adoption of public policies that guarantee access to decent work, to health programs that favor the preservation of the person's psychophysical integrity; the fight against any form of discrimination and, also, a policy that ensures broad and active participation in decisions that affect the community, especially vulnerable groups before government agencies and in decision-making processes.

Conclusions

The modernization of the Brazilian Judiciary, through automation and Artificial Intelligence, represents a significant advance in the search for greater efficiency and transparency in the provision of jurisdiction. However, as demonstrated throughout this study, this advance cannot occur at the expense of denying the centrality of the

[3] https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Accessed on 12.12.2024.

person, a fundamental principle that should guide any and all innovation in the legal field.

The use of AI in justice offers undeniable advantages, such as faster data processing and the processing of cases, but it also poses substantial risks. The lack of algorithmic transparency, the possibility of reproducing discriminatory biases, and the tendency to standardize decisions without considering the uniqueness of each individual are problems that can compromise the equity and justice of the system. The autonomy of magistrates, the right to be heard, and the possibility of reviewing automated decisions are essential elements that cannot be neglected in the implementation of these technologies.

In addition, the survey showed that the digital divide continues to be one of the biggest challenges in the digitalization of the Judiciary. Unequal access to information technologies prevents many people from fully exercising their right to defense and participating in judicial proceedings in a fair and equitable manner. Socioeconomic inequality, coupled with the lack of technological infrastructure and digital training, directly impacts vulnerable populations, further widening the gap between those who can access electronic services and those who remain excluded from the justice system.

Given this scenario, it is essential that the Judiciary adopts digital inclusion policies that ensure universal access to justice. This includes making alternative channels available to citizens without access to technology, offering digital training, and adapting electronic platforms to meet the needs of different audiences. In addition, the regulation of the use of AI in the Judiciary must ensure that its application respects fundamental rights, prioritizing human oversight, the transparency of algorithms, and the possibility of challenging automated decisions.

Finally, it is concluded that the automation of the Judiciary must be implemented with responsibility and commitment to constitutional principles. Technology should be used as a tool to strengthen human rights and social inclusion, observe the centrality of the person and not as a mechanism that reinforces inequalities and dehumanizes the judicial process. The balance between innovation and respect for human dignity must be the central axis of any digital transformation policy in the Judiciary, ensuring that the modernization of the system does not compromise justice, equity, and universal access to fundamental rights.

References

ANDRADE, José Carlos Vieira de. *The fundamental rights in the Portuguese Constitution of 1976, 2nd edition*, Coimbra, Livraria Almedina, 2001.

ARENDT, Hanna. *The Human Condition*, translated by Roberto Raposo, 10th edition, Rio de Janeiro, Forense Universitária, 2009.

ARONNE, Ricardo. The real rights in the constitutionalization of civil law, in *Revista Direito e Justiç*a, v. 39. n,2. jul/dez.2013.

ARROYO, César Landa. Constitución, Fundamental Rights, Artificial Intelligence and Algorithms. *Revista do Direito*, Universidade de Santa Cruz do Sul, Santa Cruz do Sul, n. 66, p. 139-158, jan./mar. 2022. Available at: https://online.unisc.br/seer/index.php/direito/index.

BARREA, Adriana; SALMORIA, Camila Henning; RODRIGUES, Francisco Luciano Lima. Artificial Intelligence – AI: Reflections on its use by the Judiciary. ReJuB - *Rev. Jud. Bras.*, Ed. Esp. Direito Digital, Brasília, p. 329-361, jul./dez. 2023.

BRAZIL. Constitution of the Federative Republic of Brazil of 1988. Available at: https://www.planalto.gov.br/ccivil_03/constituicao/ConstituicaoCompilado.htm.

BRAZIL. *Constitutional Amendment 45, of December 30, 2004.* Available at: https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc45.htm.

BRAZIL. General Data Protection Law (LGPD), Law No. 13,709, of August 14, 2018. Available at: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

BRAZIL. *Law No. 8900, of May 26, 1999*. Law that allows the parties to use a data transmission system for the practice of procedural acts. Available at: https://www.planalto.gov.br/ccivil_03/leis/l9800.htm.

BRAZIL. *Law No. 14,419, of December 19, 2006*. Law that provides for the computerization of the judicial process. Available at: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/l11419.htm.

Regional Centre for Studies for the Development of the Information Society. Available at: https://cetic.br/pt/sobre/. Regional Center for Development Studies of the United Nations Society .

Institute for Technology and Society (ITS), which was prepared by the School of International and Public Affairs (SIPA), at Columbia University, called "The Future of AI in the Brazilian Judicial System", available at: https://itsrio.org/wp-content/uploads/2020/06/SIPA-Capstone-The-Future-of-AI-in-the-Brazilian-Judicial-System-1.pdf.

EUROPEAN UNION. White Paper on Artificial Intelligence - A European Approach to Excellence and Trust. Available at: https://op.europa.eu/pt/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1. Noble, S. U. Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press, 2018.

INTERNATIONAL TELECOMMUNICATION UNION (ITU). *ITU Facts and Figures Annual Report.* United Nations Specialized Agency for Information and Communication Technologies – ICTs. Available at: https://www.itu.int/itu-d/reports/statistics/2022/11/24/ff22-internet-use/.

LINERA, Miguel Ángel Presno. *Fundamental Rights and Artificial Intelligence*. Madrid: Marcial Pons, Manuel Giménez Abad Foundation, 2022.

LUCAS, Clarinda Rodrigues. Information technologies and the digital divide. *Transinformação Magazine*. V. 14.n.2. p.159-165, July/December, 2002. Available at: https://periodicos.puc-campinas.edu.br/transinfo/article/view/6439.

NOBLE, S. U. *Algorithms of Oppression*: How Search Engines Reinforce Racism. NYU Press, 2018.

REGULATION (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence. Available at: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32024R1689.

RODOTÀ, Stefano. *The right to have rights*. Editora Trotta, Madrid, 2014

SILVA, Tarcízio. *Algorithmic Racism* - Artificial intelligence and discrimination in social networks. São Paulo: Edições Sesc São Paulo, 2022.

TRINDADE, Luiz P. Valerio. *It Is Not That Funny: Critical Analysis of Racial Ideologies Embedded Humour Discouses on Social Media in Brazil. 2018*. Tese (Doutorado) – University of Southampton, GB.

UN. *Universal Declaration of Human Rights*. Available at: https://brasil.un.org/pt-br/91601-declaração-universal-dos-direitos-humanos.

— 17 —

Neural data protection and the need to update the self-management model of privacy

FELIPE PAREDES
Universidad Austral de Chile
felipe.paredes@uach.cl

Introduction

In today's digital age, privacy has become a paramount concern for individuals in a scenario where information technologies are readily available and pervasive. Big tech companies gather, analyse, and trade personal data, raising alarming questions about how our information is used. Since in recent years non-invasive and wearable neurotechnologies have become increasingly accessible to the public, legislations face new dimensions of privacy challenges. These technologies, capable of collecting and processing neural data, have amplified existing concerns regarding privacy and personal autonomy.

To face threats derived from advancements in neurotechnologies, legal literature has introduced the so-called neuro-rights, a new type of human rights that includes free will, mental continuity, non-algorithmic discrimination, and equal access to mental enhancement, and, of course, mental privacy. This innovative legal framework has sparked extensive debate about the necessity of establishing new human rights in light of the evolving landscape of neuroscience and neurotechnologies. Specifically, the dialogue around brain privacy raises critical questions about whether existing privacy laws adequately could protect individuals from intrusions stemming from collecting and analysing neural data. Advocates for neuro-rights argue that the proper solution is a new mental privacy right. In contrast, opponents argue that intrusions to privacy derived from the collection and analysis of neural data are already covered by current privacy protection mechanisms.

This article takes an intermediate approach, delving into the insufficiencies of current personal data protection laws based on the privacy self-management model. This model has already been criticised as flawed in protecting regular personal data

in the present scenario. *A fortiori*, the paper emphasises that these existing laws also fail to address the complexities and sensitivities surrounding neural data.

By arguing in favour of the necessity to update current privacy protection mechanisms without incepting a new right to brain privacy is expected to help adequately protect individuals' cognitive processes. While potentially contentious in terminology, the proposal for a "human neuro-right to mental privacy" underscores the pressing need for enhanced protective measures tailored to neural data's intricacies. Ultimately, the article seeks to illuminate the critical significance of neural data and stimulate discussion on the most effective strategies for its protection in an increasingly digital world.

The paper is structured into three main sections. The first section delves into the concept of privacy, examining the mechanisms available under contemporary law to protect it and its weaknesses. The second section highlights the challenges that emerge from developing new neurotechnologies and their implications for privacy protection. Finally, the third section presents potential strategies aimed at bolstering privacy in the context of the irruption of pervasive devices processing biometric data, sketching the idea that neural data should be considered hyper-sensitive data, which reinforces safeguards for privacy in the era of portable and wearable neurotechnologies.

1. What is privacy, and how does the Law protect it?

Human rights are often regarded as fixed and immutable legal concepts, but human rights, understood as those inherent to the human condition, appear historically only in modern times, culturally linked to humanist philosophy (Habermas, 2011). The relationships between human rights and modernity are so intertwined that it is challenging to attribute contemporary meaning to the notion of human rights in previous stages of human history. Even among human rights, few legal institutions are as context dependent as the right to privacy, which has evolved in close relation to the specific context from which they emerged.

According to this, it is not surprising that the right to privacy was not included in early declarations of constitutionalism, such as the French Declaration of Rights of Men and of the Citizen or the U.S. Bill of Rights—at least not in its current form. Actually, the right to privacy was a legal creation from the 19th century, and it only emerged when specific technological innovations related to the transmission of information started to play a key role in shaping social relationships and simultaneously threatening people's basic interests.

Precisely, this dependence on contextual elements explains the plasticity of the right to privacy in terms of its evolution, to the extent that it has evolved in parallel with mass media and information technologies. As Acemoglu and Johnson (2012) note, technological development does not automatically translate into widespread prosperity since the way technologies evolve is influenced by economic, social and political decisions. Like all technology, information networks are not neutral and have also been shaped by political and economic powers. Similarly, Harari (2024) explores key dilemmas that people across different eras have faced when constructing information networks, examining how various responses to these dilemmas framed contrasting human societies. This helps to explain how the concept of privacy has evolved throughout history concerning different dimensions of the quotidian life of modern societies.

It is well known that the right to privacy is closely tied to the rise of mass media. It has been argued that the modern idea of privacy was coined in the homonymous article written by professors Warren and Brandeis, where privacy is defined as "the right to be let alone", that is, to remain outside of public scrutiny (Warren & Brandeis, 1890). According to Keulen and Kroeze (2018, p.29), journalism flourished with the advent of what has been called "New Journalism" during this period. New Journalism had a bright side since it was understood as the image of the 'modern' committed, well-informed, and respectable media who wrote columns or tried to find out what 'really' happened. However, new journalism also refers to the emergence in the U.S. of a trend called boulevardism, a style of journalism in mass media newspapers that focuses on gossip, scandal, and celebrity life.

Indeed, mass media evolved dramatically during the 20th century, first, with the invention and widespread adoption of radio and television and with the emergence of the internet lately in the last decade of the century. Additionally, companies and the government have developed increasingly sophisticated mechanisms for data collection, heightening the possibilities of profiling and tracing individuals. As information technologies changed through the years, privacy also evolved in parallel. In this light, Westin (2003) shows these developments, identifying three eras of steadily growing privacy concerns and societal responses across citizen-government, employee-employer, and consumer-business relationships in 1961–1979, 1980–1989, and 1990–2002. Each period is described in terms of new technological applications, shifting social climates, and organisational and legal developments.

As briefly outlined above, the history of privacy also reflects the difficulties in defining the concept. In general terms, the right to privacy can be defined as the right to exclude third parties from accessing certain personal information, but this definition falls short in specific contexts, and legal scholarly literature has highlighted the complexities of providing a single-dimensional definition. As put by Solove (2002, p.1094), privacy is a kind of multi-layered right and can be understood through a series of different conceptions that embody the notion of privacy within certain specific contexts: (1) the right to be left alone; (2) limited access to the self; (3) secrecy; (4) control of personal information; (5) personhood; and (6) intimacy. Furthermore, much confusion stems from terminology since the term intimacy has been preferred in Europe and Latin America rather than privacy, although the conceptual problems seem to be similar on both sides of the Atlantic (Figueroa, 2014, p. 46).

To understand how legal systems protect privacy, it is essential to take into account the dichotomy between private and public life, although this distinction is often blurred, porous, contestable, and subject to ongoing negotiation and struggle (Thomson, 2011, p. 49). To cope with this issue, legal strategies typically involve identifying various domains of protection, such as the home, family, and body, among others. These dimensions of personal life are generally understood to belong to the private sphere and, therefore, fall within the scope of privacy protection. In this light, a well-known example is the German Constitutional Court case law, which has created the notion of "spheres of protection" (Sphären-theorie). This term refers to the German approach to privacy infringements, which distinguishes between three privacy spheres—the intimate, private, and social spheres—each with a distinct protection regime (Lindemann & Van Toor, 2018, p. 377).

Under this framework, privacy is safeguarded through a series of restrictions that limit the ability to access, collect, store, manage, and transmit personal information. Privacy laws operate within a model known as "self-management of privacy." In this paradigm, privacy is an umbrella term as the law grants individuals a set of specific rights that enable them to make informed decisions about how to manage their personal data. These rights primarily include consent, access, modification, and deletion concerning the collection, use, and disclosure of personal data. These rights aim to give individuals control over their personal data, allowing them to assess the costs and benefits of its collection, use, and disclosure (Solove, 2013, p. 13). According to this model, personal data may be stored and processed based on the consent of the data subject, which can be withdrawn at any time. Additionally, administrative agencies are responsible for implementing protection mechanisms to ensure the exercise of these rights. An important distinction in this context is the difference between ordinary personal data and sensitive personal data. The latter are subject to stricter conditions, as they are closely tied to individuals' private lives.

The self-management of privacy model was initially adopted by The U.S. Privacy Act of 1974 and followed by the European Union's Data Protection Directive in 1995. Currently, this model is spread worldwide and is the predominant strategy for privacy protection. However, it was initially conceived to respond to the needs that arose in an era before the internet became ubiquitous, which has led to criticisms regarding its poor adaptation to the digital age. The proliferation of the data market on the Internet has brought to light a series of difficulties that have revealed several challenges stemming from the difficulty of controlling the information individuals provide to data market companies.

At present, companies have an enormous capacity to store and process large amounts of personal data concerning individuals' digital lives. In addition, a striving personal data market has stimulated the widespread circulation of personal information, which has become the gold of the 21st century. This is because, with the help of algorithms and artificial intelligence, accurate predictions can be made about people's preferences, interests and activities. This technology has been used for different purposes, from targeted advertising to influencing democratic processes (Véliz, 2021).

Thus, the weaknesses of the model increase their effects in the era of big data, where the information market has flourished as one of the most profitable industries and big tech companies have accumulated great power and influence worldwide (Baruh & Popescu). Solove's account has been highly influential, and recently, several studies have dug into the model's flaws, also called "notice and consent". For instance, Kröger et al. (2021) developed a long list of obstacles, rendering this strategy useless in the big data era. Among them can be mentioned: time constraints, lack of knowledge, information overload, lack of transparency, financial incentives, non-negotiability of terms, uniformity of practices, etc.

2. Personal data and neurotechnologies.

There is a vast bibliography that addresses the legal implications of neurosciences. For more than three decades, this literature has reflected on the possibility and consequences of scrutinising what occurs inside mental processes in areas of the Law

where legal categories have been built on the notion of free will, like criminal law and tort law. Seminal contributions to this discussion can be traced back to influential works such as Denno's 1988 analysis, which initiated a dialogue about the relevance of neuroscientific insights in criminal responsibility. Similarly, Restak et al. (1996) expanded on the implications of brain science for mental health and competency assessments within legal contexts. Kulynych's work in 1997 further delved into the nuances of how cognitive neuroscience could reshape our understanding of intent and agency in legal scenarios.

Currently, the significant advancements in neuroscience (NS) and neurotechnologies made in the second and third decades of the 21st century are approaching a new era in the relationship between neuroscience, neurotechnology, and the Law. The development of invasive and non-invasive neuro-technological devices has led to a rising capacity to map brain functioning and store and process information, enabling substantial progress in static and functional brain imaging (Shen, 2010). Today, it is already possible to know in detail not only the anatomy of the brain but also how this organ works internally through neural connections. While research was previously limited to static images, recent advances have allowed neuroscientists to identify specific brain functions and explore complex neural networks. As a result, imaging techniques more accurately incorporate dynamic dimensions every time, providing an initial understanding of how the human mind operates (Bareither, 2014, p. 41).

During the Obama administration was established in the U.S. the Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) aimed to expand the understanding of the human brain. By accelerating the development and application of innovative technologies, researchers have worked on developing a revolutionary new dynamic image of the brain that, for the first time, would show how individual cells and complex neural circuits interact both in time and space. This knowledge has been sought for decades because it is essential in the search for new ways to treat and even prevent brain disorders such as Alzheimer's and Parkinson's diseases. This image supposedly will fill major gaps in our current knowledge of the brain and will provide unprecedented opportunities to explore precisely how this organ registers, processes, uses and stores information. The BRAIN project has made enormous progress that allows us to glimpse the creation of increasingly precise ways of recording human brain activity in real-time, at the same time that it has begun to experiment with forms of stimulation and manipulation, such as the use of Transcranial Magnetic Stimulation, Deep Brain Stimulation and Brain-Machine Interfaces, which raises alarms about how this outlook could affect human cognitive capacities (Paredes & Quiroz, 2022).

Although the development of these neurotechnologies may take years or even decades to perfect, some commentators have noted that we are progressing toward creating devices capable of decoding people's mental processes and directly influencing the brain mechanisms that underlie what we refer to as individuals' intentions, emotions, and decisions (Yuste & Goering, 2017, p. 160). This issue gained momentum in 2017 when discussions about so-called neuro-rights emerged in legal academic journals. The neuro-rights debate stems from several important ethical concerns related to advancements in neurotechnology that the legal system needs to address: the protection of privacy concerning brain information, the potential impact

on individuals' moral agency, the possibility of enhancing cognitive abilities, and safeguarding against biases that may arise from the use of these technologies (Yuste & Goering, 2017).

The neuro-rights approach claims the need to recognise a new human rights category to face those concerns in legal terms since classical human rights could be overwhelmed by scientific and technological advances. For example, the neuro-rights proposed by Yuste *et al*. (2021, p.160) include: (1) the right to identity, or the ability to control the physical and mental integrity of people; (2) the right to agency, or freedom of thought and free will to choose one's own actions; (3) the right to mental privacy, or the ability to keep thoughts protected from disclosure; (4) the right to fair access to mental augmentation, or the ability to ensure that the benefits of improvements in mental and sensory capacity through neurotechnology are distributed fairly across the population; and (5) the right to protection against algorithmic bias, or the ability to ensure that technologies do not introduce prejudice. These proposed new human rights should be integrated into major human rights legal frameworks, such as the Universal Declaration of Human Rights.

It is still unclear how far these NTs can develop and whether, at some point, they could thoroughly do something like "reading the mind" in all its complexity or even more influence or modify thoughts. However, one of the first visible manifestations is privacy concerns. For Farahany (2023), certain forms of intrusion into the human mind, the product of emerging technologies, are so close to materialising that they impose the need for an urgent public discussion that allows the Law to develop safeguards for cognitive freedom and free will, as they would for any other area of personal freedom. According to that, some recent studies have shown their potential and imminence. For example, scientists at the University of Texas have developed a technique capable of translating people's brain activity (including unexpressed thoughts) into lifelike speech. (Tang *et al*., 2023). This non-invasive brain-computer interface (BCI) can decode continuous thoughts, combining functional magnetic resonance imaging scans and artificial intelligence language models.

Researchers had already demonstrated a few years ago that they could decode unspoken language by implanting electrodes in the brain and using an algorithm that translates brain activity into language (Moses *et al*., 2019). However, this technique requires surgery, which is why it is eventually attractive only for patients with severe pathologies for whom the benefits outweigh the costs and risks. On the other hand, until now, BCIs were effective in decoding only basic brain states such as fatigue or anger, but the University of Texas study shows the potential of a non-invasive brain-computer interface (BCI) that can decode the brain's continuous language in a way that another person could read the essence of what the patient is thinking, even without saying a single word.

The previous section highlighted the limitations of the current privacy management model, emphasising that it is outdated in today's digital landscape. Originally designed to address privacy intrusions stemming from traditional media—such as print publications, radio broadcasts, and television programming, this model is ill-equipped to handle the complexities of contemporary privacy challenges. These traditional media violations often involved invading an individual's privacy to harvest highly valuable information, primarily for commercial gain and profitability. The

present-day scenario resembles an institutional design where thieves break into jewellery shops only for targeting the most expensive diamonds, illustrating the model's focus on mainly securing and protecting what traditional media once considered high-value information.

In today's digital landscape, personal data companies generally do not distinguish between different types of information. They operate using an accumulation model that effectively merges and analyses both sensitive and non-sensitive data alongside artificial intelligence technologies. This *modus operandi* carries significant implications for privacy and data ethics as these companies can create detailed profiles that enhance their ability to predict and influence individual behaviours. The more data the model processes, the more refined its predictive capabilities become, leading to a deeper understanding of consumer habits and preferences.

The vast amount of information companies can currently collect can lead to invasive profiling—where they can seemingly predict people's actions and decisions based on their digital footprints—today's focus primarily remains on observing external behaviours. The risk associated with new technologies lies in the boundary between external observation of actions and invasive intrusion into people's thoughts and personal beliefs. Crossing this boundary could represent a significant turning point in the ongoing struggle to protect individual freedoms and privacy in an age where data-driven insights are becoming increasingly powerful.

3. Neural data as especially sensitive biometric data

According to Dantcheva *et al*. (2015), biometrics is the science of recognising individuals based on their physical, behavioural, and physiological attributes such as fingerprint, face, iris, gait and voice. Databases have increasingly started to incorporate biometric data, which includes information related to the physiological functioning of the human body. This data can encompass a variety of metrics, such as systems that monitor heart rates, fingerprints, facial recognition, and even more intricate measures like eye movement tracking or brainwave activity. The integration of biometrical data into databases is significant for several reasons. Biometrical data enables a more in-depth and detailed examination of individual behaviours by comparing biometric reactions with behavioural patterns identified through different methods. For example, researchers and analysts can explore how particular types of content consumption—like articles or videos—relate to physiological responses, including variations in adrenaline or dopamine levels, heart rate changes, or other emotional and physical involvement indicators.

Combining biometrical data with data obtained from digital footprints amplifies the scope and precision of data companies in profiling users. Crossing information using algorithms opens pathways to deepen understanding human behaviour's intricate dynamics. By analysing, for instance, the timing of these interactions, alongside biometric responses, makes it possible to draw conclusions about how specific stimuli affect individuals on a biological level (e.g. moments of heightened arousal or relaxation in response to particular types of news or social interactions).

As put by Pauzauskie *et al*. (2024), in recent years, we have seen the proliferation of consumer neurotechnology devices collecting neural data with promises to

enhance cognition, improve sleep, deepen meditative states, and balance mood, among other indications. Furthermore, companies such as Meta, Apple, Snap, and Neuralink have patented or are developing wearable neurotechnologies that will soon enter the market at world population levels never seen before. Overall, the infusion of biometric data into analytical frameworks has the potential to revolutionise our understanding of human behaviour. However, simultaneously, it raises important questions about privacy and the ethical use of such data in decision-making processes.

Of course, attempting to read human mind is not something new. Behavioural sciences have used traditional observation methods for centuries with this purpose. In the digital age, people's minds are analysed processing the traces left behind when using digital platforms, aided by powerful algorithms. With the rise of neurotechnologies, a new and more accurate way of reading people's minds is emerging through the combination of neural data scanning and all the already existing techniques. While it has not yet been proven that complex and abstract mental states can be deciphered only using neural data, we know that this data holds significant potential for interpreting and enhancing the meaning of other information easily available, functioning as a type of metadata. Consequently, there is a pressing need for regulation to protect neural data due to its unique implications.

Legislations worldwide have generally used the concept of sensitive data to establish a special regime for protecting specific personal information that requires a higher level of confidentiality due to its critical nature. However, there is no universally accepted definition of sensitive data, and many legislations avoid providing a strict definition. Instead, they present a list of types of information deemed sensitive and deserving of special protection. This list varies across jurisdictions, but common examples include ethnic origin, political opinions, religious or philosophical beliefs, and health or sexual life, among others. Quinn and Maglieri (2021, p. 1584) have observed that, like all institutions related to privacy, the concept of sensitive data has been evolving both in legal terms (de jure) and in practice (de facto), resulting in an expanded scope that now includes new categories. In any case, biometric data is always classified as sensitive data.

To understand the importance of neural data, i.e. data collected from neural activity, it is essential to comprehend that not all biometric data carries equal potential for extracting meaningful insights. Among different types of biometrical data, neural data likely has the greatest capacity to disclose critical information about an individual. Indeed, scholarly literature has distinguished between soft and hard biometrical data (Reid et al., 2013), and of course, neural data represents one clear case of hard biometric data. Revealing data from neural activity in a person would uncover myriad information about their mental states, thoughts, emotions, and behaviours (Tong & Pratte, 2012). This process, often referred to as "brain reading" or "neural decoding," involves analysing patterns of brain activity to infer what a person is experiencing or thinking. For example, studies have shown that it would be possible to decode a person's conscious experiences, such as visual perceptions and covert attitudes, based on non-invasive measurements of their brain activity (Haynes & Rees, 2006).

Arriving at this point is where it is necessary to retake the debate on neuro-rights as a potential new category of human rights. As previously mentioned, the right to privacy has become a central focus among the various proposals on neuro-rights.

Accordingly, a body of scholarship has concentrated on the need for new human rights to address the challenges posed by neurotechnological advancements, arguing the limitations of existing legal frameworks. The Neuro-rights Foundation and prominent scholars like Ienca and Andorno (2017) and Yuste and Goering (2021) have championed this perspective. In this way, the neuro-rights approach has proposed a new human right called the right to mental privacy, which seeks to protect neural data from the risks posed by emerging technologies. However, the idea of a new human right to neural privacy immediately has raised the question of whether this new right would substantively differ from the traditional notion of privacy. Most of the literature has argued that the answers to this question should be negative, claiming that such neuro-rights would be redundant (Zúñiga-Fajuri et al., 2021), even if there is little effort to determine how current legal structures can face the new challenges posed by neurotechnologies.

Previous sections of this work have provided categorical evidence demonstrating that the current privacy protection frameworks are insufficient in the era of neurotechnologies, especially considering the self-management model of privacy. Therefore, there are compelling reasons to reassess privacy protection laws to respond to neurotechnology developments, adopting an intermediate strategy in the way that Istace (2025) calls a "derived approach." It does not necessarily mean avoiding discussing whether new rights are necessary. The problem is that while this discussion is academically engaging, it may unintentionally slow down responses to pressing concerns since human rights is an arena crossed by complex philosophical and legal discussions and debates in this area could take a long time to solve. In other words, there are pragmatic reasons for adopting an adaptive model that leverages existing structures to enhance institutions to protect the right to privacy as quickly as possible. Meanwhile, scholars can continue discussing if new rights are necessary. However, there is no reason to stop legal development because, given the relevance of the issue, legislation must be able to provide suitable responses, even though they can be understood as provisory. Under these conditions, it is possible to consider brain privacy as one of the distinctive dimensions that shape the whole concept of privacy.

As explained earlier, there is nothing strange to the genetics of the right to privacy as it is a multidimensional, constantly evolving right. This would make it possible to conceive how existing mechanisms can be fine-tuned, crafting a new dimension of privacy regarding neural information. The proposal here sketched considers the auto-manage model of privacy as being overcome concerning neural data and the need for new tools. This new dimension should be protected under the category of hyper-sensitive information. In today's world, there is a growing concern about the widespread availability of non-invasive devices that are easily accessible to the public. These devices pose significant privacy challenges, such as insecure data-sharing channels, unclear and often misleading privacy policies, and an increased risk of malicious cyberattacks. As a result, stricter consent regulations for brain data are essential, and these regulations should be even more stringent for neural data due to its specific vulnerabilities and the serious consequences of its misuse. This urgent need for enhanced regulatory measures highlights the importance of protecting individuals' neural privacy in an era where technological advancements continually blur the lines between accessibility and security.

Neural data should be regarded as highly sensitive, necessitating real-time monitoring of its usage and implementing stricter requirements. The following safeguards could be discussed to ensure its protection: a. encryption, where sensitive data is encrypted both in transit and at rest to prevent unauthorised access; b. access control, which allows only authorised personnel to access the data; c. data minimisation, which involves collecting only the minimum amount of data necessary for the intended purpose to reduce the risk of exposure; d. regular audits to identify and mitigate potential vulnerabilities; and e., data breach protocols, which establish security measures for responding to data breaches, including notification procedures and remediation steps, should be in place. By following these guidelines, we can better protect neural data from unauthorised access and potential risks.

References

Bareither, I. (2014). Historia de las imágenes cerebrales. *Mente y Cerebro*, nov/dic, p. 40-48.

Baruh, L., & Popescu, M. (2017). Big data analytics and the limits of privacy self-management. *New media & society*, v. 19, n. 4, 5, p. 79-596.

Dantcheva, A., Elia, P., & Ross, A. (2015). What else does your biometric data reveal? A survey on soft biometrics. *IEEE Transactions on Information Forensics and Security*, v. 11, n. 3, p. 441-467.

Denno, D. W. (1988). Human biology and criminal responsibility: Free will of free ride. *University of Pennsylvannia Law Review,* v. 137, p. 615–671.

Farahany, N. A. (2023). *The Battle for Your Brain*: Defending the Right to Think Freely in the Age of Neurotechnology. St. Martin's Press.

Figueroa, R. (2014). *Privacidad*. Santiago de Chile: Ediciones Universidad Diego Portales.

Habermas, J. (2011). The Concept of Human Dignity and the Realistic Utopia of Human Rights. In *Philosophical Dimensions of Human Rights: Some Contemporary Views*, Dordrecht: Springer Netherlands, p. 63–79.

Harari, Y. N. (2024). *Nexus*. A Brief History of Information Networks from the Stone Age to IA. Random House.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature reviews neuroscience*, v. 7, n. 7, p. 523–534.

Ienca, M. y Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology, *Life Sciences, Society and Policy*, v. 13, n. 5.

Istace, T. (2025). Establishing Neurorights: New Rights versus Derived Rights. *Journal of Human Rights Practice*, v. 17, n. 1, p. 121–139.

Johnson, S., & Acemoglu, D. (2023). *Power and progress*: Our thousand-year struggle over technology and prosperity. Hachette UK.

Keulen, S., & Kroeze, R. (2018). Privacy from a historical perspective. *The handbook of privacy studies:* An interdisciplinary introduction, 21-56.

Kröger, J. L., Lutz, O. H. M., & Ullrich, S. (2021). *The myth of individual control*: Mapping the limitations of privacy self-management.

Kulynych, J. (1997). *Psychiatric neuroimaging evidence*: A high-tech crystal ball? *Stanford Law Review*, p. 1249–1270.

Lindemann, M., & Van Toor, D. A. G. (2018). Protection of a Suspect's Privacy in Criminal Procedures. Does the Conceptual Approach of the German Federal Constitutional Court Make a Difference? *Ars Aequi Mei*, p. 376-384.

Magee, P., Ienca, M., & Farahany, N. (2024). *Beyond neural data*: Cognitive biometrics and mental privacy. *Neuron*, v. 112, n. 18, p. 3017-3028.

Moses, D.A., Leonard, M.K., Makin, J.G. *et al.* (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nat Commun*, v. 10, n. 3096.

Paredes, F., & Quiroz, C. (2022). Neuroderechos en Chile: Estado del arte y desafíos. In: Lopes, Ana Maria D'Ávila, A. *et al. Neurodireito, Neurotecnologia e Direitos Humanos*. Porto Alegre: Livraria do Advogado, 69-80.

Pauzauskie, S., Genser, J., & Yuste, R. (2024). Protecting Neural Data Privacy—First, Do No Harm. *JAMA neurology*.

Quinn, P., & Malgieri, G. (2021). The difficulty of defining sensitive data—The concept of sensitive data in the EU data protection framework. *German Law Journal*, v. 22, n. 8, p. 1583-1612.

Reid, D. A., Nixon, M. S., & Stevenage, S. V. (2013). Soft biometrics; human identification using comparative descriptions. *IEEE Transactions on pattern analysis and machine intelligence*, v. 36. n. 6, p. 1216–1228.

Restak, R. M. (1996). Brain Damage and Legal Responsibility. In *Seminars in Clinical Neuropsychiatry*, v. 1, n. 3, p. 170-171.

Shen, F (2010). The Law and Neuroscience Bibliography: Navigating the Emerging Field of Neurolaw, *International Journal of Legal Information*, v. 38, n. 3, p. 352–399.

Solove, D. J. (2002). Conceptualizing privacy. *Calif. L. Rev.*, *90*, 1087.

Solove, D. J. (2013). Autogestión de la privacidad y el dilema del consentimiento. *Revista Chilena de Derecho y Tecnología*, v. 2, n. 2, p. 1.

Tang, J., LeBel, A., Jain, S. *et al.* (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat Neurosci*, v. 26, p. 858–866.

Thompson, J. B. (2011). Shifting boundaries of public and private life. *Theory, culture & society*, v. 28, n. 4, p. 49–70.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual review of psychology*, v. 63, n. 1, p. 483-509.

Véliz, C. (2021*). Privacy is Power*: Why and How You Should Take Back Control of Your Data. Melville House Printing.

Warren S.M. & Brandeis, L. (1890). The Right to Privacy. *Harvard Law Review*, v. 4, n. 193.

Westin, A. F. (2003). Social and political dimensions of privacy. *Journal of Social Issues*, v. 5, n. 9, p. 431-453.

Yuste, R. y Goering, S. (2017). Four Ethical Priorities for Neurotechnologies and AI, *Nature*, v. 551, p. 159-163.

Yuste R., Genser, J. y Herrmann, S. (2021). It's Time for Neuro-Rights, *Horizons,* v. 18: p. 154-164.

Zúñiga-Fajuri, A., Miranda, L. V., Miralles, D. Z., & Venegas, R. S. (2021). Neurorights in Chile: Between neuroscience and legal science. *Developments in neuroethics and bioethics*, v. 4, p. 165-179.

— 18 —

The commodification of free will: neurotechnology and ethical challenges in the workplace

ALINE PASSOS MAIA
Universidade de Fortaleza
alinepmaia@unifor.br

RAQUEL PASSOS MAIA
Universidade de Lisboa
raquelpmaia13@gmail.com

LARISSA MACIEL DO AMARAL
Universidade de Fortaleza.
larissaamaral@unifor.br

Introduction

Neurotechnology is reshaping workplaces, introducing opportunities for innovation while raising significant ethical and legal concerns. This new frontier allows for the collection and analysis of neural data, encompassing cognitive states, emotional responses, and even thought patterns. While enthusiasts emphasize the potential for increased productivity and enhanced employee well-being, critics warn of the risks to mental privacy, autonomy, and dignity.

The commodification of neural data introduces profound ethical dilemmas. For instance, can employers claim ownership over the cognitive processes of their workforce? Or should such data be protected as an intrinsic part of individual autonomy? These questions challenge traditional legal notions of property and privacy, urging the development of robust regulatory frameworks. This chapter explores these com-

plex issues, addressing the balance between innovation and a new pattern of fundamental human rights in the workplace.

1. Neurotechnology and the capital of surveillance

In the workplace, neurotechnology commodifies neural data, transforming cognition into a tradable asset under the framework of surveillance capitalism. EEG monitoring and emotional analysis grant employers deep insights but raise ethical concerns, including discrimination, pressure to optimize, and loss of privacy. This risks workplace inequalities and challenges labor rights, as workers may be coerced into cognitive surveillance. Legally, existing frameworks struggle to protect mental autonomy. Ethically, it reduces individuals to data points, undermining human dignity. Urgent regulations are needed to prevent exploitation and safeguard cognitive freedom.

1.1. The commodification of neural data

The emergence of neurotechnology has ushered in an unprecedented era where human cognition itself becomes a measurable and tradable commodity. Neural data—encompassing brainwave patterns, emotional responses, and cognitive states—have been elevated to the status of "capital neural," serving as a critical asset in modern workplaces. Employers, through wearable devices and neurotech tools, now seek insights into workers' mental processes, from engagement levels to stress thresholds.

The commodification of workers' neural data refers to the transformation of cognitive processes—such as attention, emotions, and decision-making—into economic assets that can be collected, analyzed, and monetized by employers. By turning brain activity into a commercial asset, the boundaries between personal thought and corporate control become increasingly blurred, necessitating robust regulatory protections.

It's possible to mention the Emotiv MN8, a neurotech device marketed as a productivity-enhancing tool, monitors employees' focus and stress levels in real time. While enthusiasts argue that it promotes workplace well-being, critics view such practices as invasive, stripping workers of their cognitive privacy.

This commodification introduces ethical dilemmas: Do employers have the right to claim ownership over their employees' mental states? Or should neural data be protected as an intrinsic part of individual autonomy? Such questions challenge traditional legal notions of property and privacy, urging regulatory frameworks to address this unique intersection of technology and labor.

1.2. The role of technology in shaping employment dynamics and workplace inequality

Neurotechnology aligns with the broader context of "surveillance capitalism", a term coined by Shoshana Zuboff to describe the economic model where data extraction fuels profit. In this paradigm, neurotechnology extends the frontier of surveillance from the physical and digital spheres into the cognitive domain (Zuboff, 2019).

The mechanisms for data collection in neurotechnology primarily include three main approaches. First, brainwave monitoring measures focus, engagement, and cog-

nitive load through EEG-based tools such as Emotiv MN8 and Neurosky Mindwave (Maia *et al*., 2022). Second, emotional response analysis tracks emotional states using brain signals or physiological responses like heart rate, with Affectiva Emotion AI being a notable example. Lastly, neural profiling uses cognitive patterns to assess compatibility with specific tasks or roles, as seen in systems like HireVue AI Recruitment (Paredes *et al*., 2023).

By leveraging these tools, employers gain unprecedented insights into their workforce. However, these practices raise ethical concerns, particularly in scenarios where refusing to participate may result in adverse career consequences. A notable example occurred in China, where the Hangzhou Zhongheng Electric company implemented EEG-based "smart caps" to monitor workers concentration during shifts (Gómez, 2022). Although the data collected was used to optimize workflows, it also generated fears of punitive measures against employees deemed less "focused", blurring the line between efficiency and exploitation.

Furthermore, the integration of neurotechnology in labor relations risks exacerbating preexisting inequalities. Workers with cognitive profiles deemed "suboptimal" could face discrimination, while those with highly adaptive neural traits might receive preferential treatment. These impacts can be categorized into three main areas:

a) Discrimination

The exclusion of candidates based on "neural inefficiencies" represents a profound threat to workplace equity. Neurotechnology allows employers to identify and categorize workers by their cognitive profiles, potentially leading to hiring practices that prioritize individuals with higher neural focus or productivity levels. This creates an artificial hierarchy where individuals who do not align with these standards—whether due to neurological diversity, temporary stress, or other factors—are unfairly marginalized. For example, neuroatypical individuals, such as those with ADHD or anxiety disorders, may face systematic exclusion from certain roles despite their potential to excel with appropriate accommodations.

Moreover, the reliance on neural data for decision-making can reinforce existing biases, as algorithms trained on biased datasets might inadvertently perpetuate systemic inequalities based on gender, race, or socioeconomic status. This raises urgent ethical and legal questions about how to ensure fair hiring practices and protect diverse cognitive profiles in the workforce.

b) Increased pressure to "optimize"

The expectation for employees to use neuro-enhancement tools to remain competitive represents a shift in workplace norms that can have detrimental effects on individual well-being and organizational culture. As neurotechnology becomes more accessible, workers may feel compelled to adopt tools such as cognitive enhancers, brain stimulation devices, or mindfulness apps to boost productivity or focus. While these technologies promise benefits, they also introduce significant risks. Continuous optimization demands can lead to burnout, as employees are pressured to constantly improve their cognitive performance to meet escalating expectations.

Furthermore, the normalization of neuro-enhancement in the workplace could create a two-tiered labor market, dividing those who have access to these technologies from those who do not, further exacerbating existing inequalities. For example,

workers from disadvantaged backgrounds may lack the financial means to invest in neuro-enhancement tools, thereby limiting their opportunities for career advancement. This dynamic raises critical questions about consent, as the distinction between voluntary and coerced participation becomes increasingly blurred in highly competitive environments.

c) Loss of autonomy and privacy

Invasive neuro-monitoring technologies create a direct threat to workers' autonomy and privacy, fundamentally altering the relationship between employees and employers. The ability to continuously track cognitive states, such as focus or emotional engagement, shifts power dynamics within the workplace, granting employers unprecedented access to the inner workings of employees' minds. This can lead to scenarios where workers feel they have little control over their mental states, as they must constantly regulate their emotions and thought processes to align with organizational expectations.

For instance, in high-stress environments with continuous EEG monitoring, employees may suppress natural cognitive fluctuations out of fear that perceived lapses in concentration could result in punitive actions or dismissal. Such practices not only undermine mental autonomy but also contribute to a culture of hyper-surveillance that erodes trust between employers and employees. The commodification of cognitive states raises profound ethical questions about the limits of workplace monitoring and the extent to which an individual's mental processes should remain inviolable.

These practices present significant legal challenges. For instance, the principle of non-discrimination enshrined in many Constitutions Laws, such as Brazil's (Art. 7, XXX), may not yet adequately address biases introduced by neurotechnology. Additionally, labor laws must evolve to ensure that neural data is not weaponized to perpetuate inequality.

1.3. Ethical questions in the commodification of thought

Beyond its legal dimensions, the commodification of thought raises profound ethical concerns. Is it morally justifiable to monetize human cognition? Critics argue that such practices risk dehumanizing workers, reducing them to mere neural components optimized for productivity. This reductionist approach disregards the holistic value of human effort, creativity, and individuality—traits that transcend measurable neural outputs.

From a philosophical perspective, the commodification of thought directly challenges Kantian ethics, which posit that individuals should always be treated as ends in themselves, not merely as means to an economic goal. By extracting and monetizing cognitive data, neurotechnology risks violating this principle, as workers are transformed into data points within systems designed primarily for profit maximization. This shift not only undermines the intrinsic dignity of labor but also raises concerns about the erosion of personal identity in the face of technological standardization.

Additionally, these practices echo critiques rooted in Marxist theory, which identifies labor as a fundamental expression of human creativity and self-actualization. By converting thought into a tradable commodity, neurotechnology perpetuates the alienation of workers from their own cognitive processes, further deepening the

divide between human agency and economic utility. The worker's mental effort, once an integral part of their individuality, becomes subsumed into a broader system of surveillance capitalism.

The ethical implications of this commodification extend beyond the workplace. When cognitive data becomes an economic asset, it risks reinforcing systemic inequities, as access to and control over this data often lie with powerful corporations. This dynamic exacerbates the vulnerability of workers, particularly those in precarious or low-income positions, who may feel compelled to consent to invasive monitoring out of economic necessity. Such coercion undermines the principles of informed consent and autonomy, cornerstones of modern ethical frameworks.

Furthermore, neurotechnology's ability to reshape labor dynamics raises questions about the moral boundaries of efficiency. While increased productivity may be a laudable goal, it must not come at the cost of workers' mental and emotional well-being. The relentless pursuit of optimization through cognitive surveillance risks fostering environments of constant pressure and competition, where workers are judged not for their holistic contributions but for their neural efficiency.

These ethical dilemmas underscore the urgency of establishing robust safeguards to protect workers' autonomy and dignity. Regulatory frameworks must address not only the potential misuse of neural data but also the broader implications of commodifying thought itself. Ethical guidelines should emphasize the preservation of cognitive freedom, the right to mental privacy, and the recognition of labor as an inherently human act that cannot be reduced to neural metrics.

In synthesis, the commodification of thought through neurotechnology poses ethical challenges that strike at the core of what it means to be human. As workplaces increasingly integrate cognitive surveillance, society must confront these dilemmas with a commitment to upholding the intrinsic value of human effort, ensuring that technological advancements serve humanity rather than exploit it.

2. The new asset: neural data in the economic context

Neural data, extracted from brain activity, is becoming a key economic asset, extending surveillance capitalism into cognition. While it promises workplace optimization, it raises ethical concerns about privacy, autonomy, and discrimination. Existing laws fail to address its risks, allowing potential exploitation and deepening inequalities. Without strict regulations, neural data could dehumanize labor, reducing workers to mere productivity metrics. Safeguards must ensure transparency, protect workers' rights, and prevent coercion.

2.1. Neural data: a new economic pillar

In the 21st century, personal data has become a cornerstone of the global economy. Yet, the advent of neurotechnology is pushing this frontier further by introducing neural data—information directly extracted from brain activity. This new form of data includes metrics such as levels of attention, emotional states, and cognitive reactions, offering employers unprecedented insights into the mental states of workers.

Neural data holds immense economic value due to its real-time accuracy and depth. Companies are integrating neurotechnologies to monitor and optimize workforce productivity. However, this integration also raises significant ethical concerns, particularly when neural data is commodified without clear safeguards. Regarding this, it's important to delineate the key differences between personal and neural data.(Frame 1)

Frame 1 - Personal Data *x* Neural Data

Type of Data	Source	Primary Use	Ethical Risks
Personal Data	Social media, online transactions	Marketing, targeted advertising	Privacy breaches, data misuse
Financial Data	Banking systems, credit reports	Risk assessment, fraud prevention	Economic exclusion based on profiling
Neural Data	Brain-computer interfaces, EEGs	Workforce optimization, mental monitoring	Invasion of mental privacy, exploitation

Source: By the authors

Neural data's unique nature transforms it into a form of "capital neural," a highly sensitive asset with the potential to disrupt existing legal and ethical frameworks. The incorporation of neural data aligns with the mechanisms of surveillance capitalism, a term coined by Shoshana Zuboff to describe the commodification of human behavior through data collection. Neurotechnologies enable employers not only to analyze physical actions but to delve into cognitive patterns, further extending the scope of surveillance capitalism.

For example, tech companies like Emotiv have developed brain-monitoring devices marketed as tools for enhancing workplace efficiency. While these devices can track engagement and stress levels, they also blur the lines between innovation and intrusion. Such practices exemplify the double-edged nature of neural data in the workplace.

The current legal frameworks, such as the European GDPR and Brazil's LGPD, are insufficient to address the complexities of regulating neural data .These regulations, while robust in safeguarding traditional personal data, lack explicit provisions for the unique nature of neural information. This regulatory gap creates a vacuum that allows for potential exploitation, leaving neural data vulnerable to misuse and raising concerns about privacy and autonomy in the era of neurotechnology (Andorno, 2023).

2.2. The commodification of thought

Neural data represents a paradigm shift in how human cognition is treated economically. The commercialization of thoughts, emotions, and cognitive processes raises significant ethical and legal concerns, challenging fundamental principles such as mental autonomy and the intrinsic dignity of the individual. The act of commodifying such deeply personal data undermines the sanctity of mental autonomy, transforming what was once considered inviolable into a tradable asset. The ethical dilemmas of commodification could be schematized regarding their challenges and potential outcomes. (Frame 2)

Frame 2 - Commodification of neural data ethical challenges

Aspect	Challenge	Potential Outcome
Mental Autonomy	Loss of control over one's cognitive data	Reduced trust in workplace environments
Exploitation Risk	Monetization of neural data without consent	Marginalization of vulnerable worker populations
Dignity of the Worker	Reduction of individuals to "data points"	Dehumanization in professional relationships

Source: By the authors

The commodification of neural data fundamentally threatens workers' control over their own mental processes. Once cognitive data becomes a tradable commodity, employees may lose the ability to regulate how their thoughts and emotions are accessed, interpreted, or used by employers. This erosion of autonomy can lead to an environment of mistrust, as workers feel that their most personal and intimate experiences are being monitored and exploited. The lack of transparency in how neural data is collected and utilized exacerbates this issue, creating a sense of vulnerability and fear among employees. For instance, continuous EEG monitoring in the workplace could compel workers to self-regulate in ways that compromise their natural cognitive rhythms, leading to stress and reduced well-being.

The monetization of neural data, especially without explicit consent, introduces a significant risk of exploitation. Vulnerable worker populations, such as those in low-income roles or precarious employment, are particularly at risk, as they may feel pressured to consent to invasive practices out of economic necessity. This dynamic mirrors broader patterns of inequality, where those with fewer resources are more likely to be subject to exploitative labor practices. The absence of robust legal safeguards further compounds the problem, as it allows corporations to extract value from neural data without adequately compensating or protecting the individuals whose data is being used. This exploitation creates a dangerous precedent, reducing workers to mere sources of economic value rather than respecting their humanity and agency.

One of the most concerning aspects of neural data commodification is its potential to undermine the inherent dignity of workers. By reducing individuals to "data points," workplaces risk dehumanizing professional relationships, shifting the focus from collaboration and creativity to mere productivity metrics. Philosophical perspectives, such as those of Immanuel Kant, caution against treating humans as mere means to an economic end, emphasizing the intrinsic worth of every individual. The transformation of workers into units of data not only violates this ethical cornerstone but also damages organizational culture, fostering environments where employees feel undervalued and interchangeable. Over time, this dehumanization can lead to diminished morale, higher turnover rates, and a loss of the intangible qualities that make human labor irreplaceable—such as empathy, innovation, and adaptability.

Philosophical frameworks provide valuable insights into the risks associated with the commodification of thought. However, practical solutions are also necessary to address the ethical dilemmas posed by this paradigm shift. Robust regula-

tions must be established to ensure that neural data is collected, stored, and utilized transparently and consensually. Furthermore, ethical guidelines should reinforce the importance of preserving mental autonomy, protecting vulnerable populations, and upholding the dignity of all workers. Without such safeguards, the commodification of thought threatens to deepen existing inequalities, erode trust, and compromise the foundational principles of labor ethics.

2.3. Neural data and economic inequalities

The integration of neural data into workplace practices presents a dual-edged sword: while it offers significant opportunities for innovation and productivity, it also risks exacerbating preexisting socio-economic inequalities. Workers possessing "desirable" neural traits—such as high cognitive responsiveness or emotional stability—may gain preferential treatment, creating a hierarchy of neural privilege. Conversely, individuals with cognitive profiles deemed "suboptimal" could face exclusion or marginalization, particularly in hiring processes or performance evaluations.

For example, hiring algorithms designed to prioritize high cognitive responsiveness may unintentionally discriminate against neurodiverse candidates, such as those with ADHD, autism, or anxiety disorders. This perpetuates systemic biases and undermines workplace diversity. Furthermore, as neural data becomes a valuable economic asset, disparities in access to neurotechnology could deepen inequalities, with wealthier workers or organizations reaping disproportionate benefits from advanced tools, leaving others behind.

To address these challenges, proactive measures are essential. The following safeguards can help ensure that neural data is integrated into workplace practices ethically and equitably

a) Regulatory Reforms

Neural data should be classified as sensitive data under international privacy laws, placing it on par with biometric and genetic information. This classification would mandate stringent protections for its collection, storage, and use, ensuring that neural data is handled with the utmost care and respect for individuals' mental autonomy. For instance, international frameworks like the GDPR (General Data Protection Regulation) could serve as models, incorporating explicit provisions for the ethical handling of neural data. Such reforms would establish a baseline for accountability, limiting the scope for misuse and exploitation.

b) Worker Protections

Transparency and informed consent must be cornerstones of neural data collection and usage in the workplace. Employees should have the right to understand how their neural data will be used, who will access it, and for what purposes. This includes providing clear, accessible documentation and allowing workers to opt out of data collection without fear of retaliation. Additionally, organizations should implement robust policies to safeguard against coercion, ensuring that participation in neurotechnology initiatives is genuinely voluntary. Worker protections must also address the potential misuse of data, with clear guidelines on how neural information can and cannot influence decisions about hiring, promotions, or terminations.

c) Corporate Accountability

To deter misuse and exploitation of neural data, organizations must face tangible consequences for unethical practices. Establishing penalties for violations—such as hefty fines or legal action—can act as a strong deterrent. Additionally, companies should be required to conduct regular audits of their neurodata practices, ensuring compliance with ethical and legal standards. These audits should involve third-party oversight to maintain objectivity and transparency. Beyond punitive measures, companies should also be incentivized to adopt best practices in neurotechnology integration, such as receiving certifications for ethical handling of neural data.

The commodification of neural data, as a new form of capital, has the potential to reshape economic and legal landscapes. However, without robust safeguards, its integration risks perpetuating systemic inequalities, undermining mental autonomy, and eroding worker dignity. By implementing comprehensive regulatory reforms, prioritizing worker protections, and holding corporations accountable, society can navigate these challenges while ensuring that neurotechnology serves as a force for equity and innovation, rather than exploitation.

3. The worker as a product: invisible impacts

Neurotechnology can turn workers into products, reducing them to neural metrics for corporate efficiency while eroding mental privacy. Continuous cognitive monitoring increases burnout, distrust, and mental health disparities, especially for neurodiverse individuals. Without transparency and legal protections, employees face coercion and exploitation. Ethical safeguards must prevent invasive surveillance and prioritize worker well-being over productivity.

3.1. Mental privacy vs. corporate efficiency

The integration of neurotechnology into workplace dynamics reveals a profound conflict: the pursuit of corporate efficiency versus the preservation of mental privacy. Unlike traditional surveillance methods, neurotechnologies penetrate the cognitive realm, extracting data directly from workers' brains. This unprecedented intrusion into the private domain of thoughts and emotions not only challenges existing labor rights but also raises serious ethical concerns.

Mental privacy, often regarded as the final bastion of individual autonomy, is now under threat. Devices such as EEG-based headsets allow employers to measure focus, engagement, and stress levels in real time. Although marketed as tools to enhance productivity and employee well-being, these technologies often lack transparent consent mechanisms, effectively transforming workers into unwilling subjects of cognitive surveillance. For instance, employees may be unaware of how their neural data is being collected, stored, or used, further undermining their autonomy.

The legal frameworks currently in place struggle to keep pace with these technological advancements. While labor laws robustly protect physical privacy, they are ill-equipped to address the nuances of neural data. For example, Brazil's LGPD (Lei Geral de Proteção de Dados) and Europe's GDPR (General Data Protection Regulation) classify certain categories of personal data as sensitive but fail to explicitly

consider the invasive nature of neurotechnologies. This gap creates a regulatory vacuum where coercion and misuse thrive, leaving workers vulnerable to exploitation.

To safeguard mental privacy, it is imperative to establish legal protections that explicitly address neural data. These should include provisions for informed consent, transparency in data use, and strict limitations on the purposes for which cognitive data can be collected (Gómez, 2022). Additionally, companies must be held accountable for any breaches, with penalties sufficient to deter unethical practices.

3.2. The invisible costs to mental health

While enthusiasts of neurotechnology argue that it can reduce workplace stress and optimize workflows, the reality often paints a more complex and troubling picture. Continuous cognitive monitoring places workers under immense psychological pressure, particularly when they feel their neural states are under constant scrutiny. This phenomenon, referred to as the "optimization syndrome," exemplifies the mental toll of striving for perpetual productivity in an environment where performance is reduced to neural metrics.The psychological impacts of this relentless surveillance are profound and some of them could be highlighted above.

a) Increased Burnout

Employees subjected to constant cognitive monitoring frequently report feelings of exhaustion and alienation. The pressure to maintain "optimal" states of focus and engagement often results in physical and emotional fatigue, with workers fearing that any deviation from prescribed metrics might jeopardize their job security.

b) Erosion of trust

A lack of transparency regarding how neurotechnologies are deployed and what data is being collected undermines trust between employers and employees. Workers who feel surveilled are less likely to perceive their workplace as a safe or supportive environment, leading to decreased morale and productivity.

c) Mental Health Disparities

The expectation to achieve and maintain optimal cognitive states disproportionately affects neurodiverse individuals or those with mental health conditions. For these workers, the stress of conforming to rigid neural standards exacerbates inequalities, further marginalizing already vulnerable groups.

Employers must strike a careful balance between leveraging neurotechnology for productivity gains and protecting employees' mental health. This involves adopting ethical practices that prioritize worker well-being over short-term efficiency. International labor organizations, such as the International Labour Organization (ILO), emphasize the importance of mental health as a core component of workplace safety and productivity. Companies should align with these principles by: implementing policies that limit the use of neurotechnology to non-invasive, consensual applications; providing access to mental health resources for employees subjected to cognitive monitoring; ensuring transparency in how neurotechnologies are deployed and how neural data is handled.

By addressing these invisible costs, employers can foster an environment that values both innovation and human dignity, paving the way for a future where neurotechnology enhances, rather than diminishes, workplace dynamics.

4. The mind as an inviolable space

Cognitive freedom is a fundamental right threatened by neurotechnology, which enables external influence over thoughts and decisions. Workplace neuro-monitoring risks undermining free will, turning workers into tools for corporate goals. Without strict regulations, neural data misuse can erode autonomy, blur moral responsibility, and commodify thought. Countries like Chile and Brazil are pioneering legal protections, but global standards are urgently needed to safeguard mental privacy and human dignity.

4.1. Cognitive freedom as a fundamental right

The quick advancement of neurotechnology has brought new urgency to the debate over cognitive freedom—the right to think, reason, and make decisions free from external interference or manipulation. While this right is often considered implicit within broader human rights frameworks, the growing capabilities of neurotechnologies demand its explicit recognition and protection. It's necessary to discuss the Key Dimensions of Cognitive Freedom like freedom, autonomy and privacy.

a) Freedom of Thought

This dimension safeguards individuals from intrusive monitoring or alteration of their mental processes. In a world where neurotechnology enables real-time tracking of focus and emotional states, the right to freely think without surveillance or external influence is paramount.

b) Autonomy of Decision-Making

Cognitive freedom entails protecting individuals' capacity to act based on their own reasoning, unaffected by technological manipulation. This is particularly critical in contexts where neurotechnologies might be used to subtly influence choices, whether in consumer behavior, political preferences, or workplace performance.

c) Mental Privacy

As perhaps the most fundamental aspect of cognitive freedom, mental privacy ensures the right to keep one's thoughts and cognitive states private. This principle extends beyond protection from employer or corporate intrusion to include safeguarding against state actors' potential misuse of neurotechnologies for surveillance or control.

Chile has emerged as a global pioneer in addressing these concerns through legislation. In 2021, it became the first country to explicitly enshrine neuro-rights in its constitution. This groundbreaking legal framework recognizes mental privacy and cognitive liberty as fundamental rights, setting a precedent for the protection of individuals against the misuse of neurotechnologies. The legislation mandates that neural data be treated as sensitive personal data and imposes strict limitations on its collection and use. While Chile has led the charge, other countries, including Brazil, are beginning to engage with the concept of cognitive freedom and neuro-rights.

In 2023, Brazil advanced discussions on cognitive freedom through Proposta de Emenda Constitucional n.º 29/2023. This proposal aims to amend Article 5 of the Federal Constitution to include the protection of mental integrity and algorithmic transparency as fundamental rights. The proposed text, which would be added as

item LXXX, states: "LXXX - scientific and technological development will ensure mental integrity and algorithmic transparency, in accordance with the law"

If approved, this amendment would position Brazil as a leader in safeguarding cognitive autonomy and regulating technological advancements, aligning with global trends while addressing the unique challenges posed by neurotechnology in the national context.

In the European context, Spain has initiated discussions on integrating neuro-rights into its legal framework. Proposals include treating neural data as highly sensitive information under the General Data Protection Regulation (GDPR), with additional layers of protection against its misuse.

Although no federal legislation explicitly addresses neuro-rights, certain states have taken steps in this direction. For example, California's Consumer Privacy Act (CCPA) includes provisions that could potentially cover neural data under its broader privacy protections. Furthermore, ethical debates on neurotechnology are gaining traction in academic and policy circles, signaling a growing awareness of the need for regulation.

At the supranational level, the EU is exploring regulatory pathways to address emerging neurotechnologies. The European Parliament has called for robust ethical guidelines and legal frameworks to ensure that neural data is collected and used responsibly, particularly in employment and healthcare contexts.

Australia and New Zealand are actively monitoring neurotechnology's implications through research initiatives and public consultations. Discussions focus on ensuring that advancements in neurotechnology do not compromise individual freedoms or exacerbate social inequalities.

The explicit recognition of cognitive freedom in legal frameworks is not merely a legal necessity but a moral imperative. Chile's neuro-rights legislation and Brazil's constitutional amendment proposal serve as models for global adoption, demonstrating how nations can proactively address the ethical challenges posed by neurotechnology. However, a truly effective response requires international cooperation, with nations adopting shared principles and standards to prevent exploitation and ensure that neurotechnology enhances human flourishing rather than undermines it.

4.2. Free will and the mind as a sanctuary: protecting cognitive autonomy in the age of neurotechnology

Neurotechnology brings renewed scrutiny to age-old debates about free will. By unveiling the neural mechanisms underlying decision-making, these technologies challenge traditional notions of human agency. If thoughts and actions can be influenced, predicted, or even manipulated through neurodata, the philosophical foundation of free will becomes increasingly tenuous, raising profound ethical, legal, and societal questions.

Consider, for instance, a workplace where neurofeedback tools subtly guide employees' decision-making to align with corporate goals. While this may enhance productivity, it fundamentally alters the nature of individual choice. Are these decisions genuinely the workers' own, or are they the product of engineered cognitive states designed to serve organizational objectives? The implications of such scenarios

extend beyond efficiency: they challenge the authenticity of human agency and the moral integrity of decision-making processes.

Moreover, if neurotechnology can shape decisions, the attribution of moral and legal responsibility becomes significantly more complex. Workers whose actions are influenced by cognitive manipulation may not fully own their choices, necessitating a reevaluation of foundational concepts such as liability, intent, and consent. This raises critical questions for legal systems and ethical frameworks, which must adapt to the nuances introduced by neurotechnological interventions.

In light of these challenges, the mind must be regarded as a sanctuary—an inviolable space immune to commodification or intrusion. This perspective aligns with international human rights principles that prioritize individual dignity, autonomy, and freedom of thought. Safeguarding cognitive autonomy is essential not only for preserving individual rights but also for ensuring that neurotechnological advancements do not erode the foundations of human agency and responsibility.

Addressing the ethical and practical implications of neurotechnology requires a multi-faceted approach that combines legal, corporate, and societal initiatives. First, the establishment of international standards is essential to recognize cognitive freedom as a fundamental right. These frameworks should explicitly define mental privacy and autonomy, ensuring that neural data is treated as sensitive information subject to stringent protections. By adopting global agreements, nations can harmonize efforts to regulate the ethical use of neurotechnologies and prevent their misuse.

Second, corporations must embrace transparency and accountability in their deployment of neurotechnologies. Policies should mandate clear communication about how these tools are used, what data is collected, and how it is stored or shared. Additionally, limitations must be enforced to prevent neurotechnologies from manipulating workers' cognitive states beyond acceptable ethical boundaries. Corporate responsibility should extend to fostering environments where workers feel empowered to make autonomous decisions, free from undue influence or coercion.

Finally, public awareness campaigns are crucial to educate workers and the general population about the risks and rights associated with neurotechnology. Empowering individuals with knowledge about their cognitive freedoms and the ethical dimensions of neural data use will foster informed participation in the evolving landscape of neurotechnological applications. Such campaigns should emphasize the importance of mental privacy, the risks of commodifying thought, and the value of preserving the sanctity of the mind.

The philosophical and practical challenges posed by neurotechnology underscore the urgent need to safeguard the sanctity of the human mind. Cognitive freedom is not merely an abstract ideal but a cornerstone of individual dignity and autonomy. As neurotechnological advancements continue to unfold, maintaining the mind as an inviolable space will be critical to balancing the benefits of innovation with the preservation of humanity's most fundamental rights.

Conclusion

The rise of neurotechnology marks the beginning of a transformative era, brimming with opportunities and challenges, particularly in the workplace. Neural data,

with its unprecedented depth and precision, offers the potential to revolutionize industries, enhance productivity, and unlock new dimensions of human potential. Yet, this same technology poses existential threats to fundamental rights, such as mental privacy, cognitive autonomy, and human dignity.

The integration of neurotechnology into workplace practices requires a delicate balancing act. While its promise of increased efficiency and improved well-being is enticing, the risks of exploitation, inequality, and commodification of the mind demand urgent attention. As this article has explored, the ethical and legal dilemmas arising from the commodification of thought are profound and cannot be ignored.

Key questions remain at the forefront of this technological revolution: How can we protect workers' rights in an era where thoughts can be monitored and monetized? What safeguards are necessary to ensure that neurotechnology serves humanity rather than exploits it? Where should we draw the line between innovation and intrusion into the sanctity of the human mind?

The answers to these questions lie in a multifaceted approach grounded in proactive regulation, ethical vigilance, and global cooperation. Legal frameworks must evolve to address the unique challenges posed by neurotechnology. Laws should incorporate core principles such as informed consent, ensuring that workers have full control over the use of their neural data and can opt out without fear of retaliation. Transparency and accountability are equally essential; organizations must disclose how neurotechnologies are implemented, what data is collected, and how it is used, with clear penalties for misuse. Regulatory frameworks must also harmonize globally to prevent exploitation and ensure that neurotechnology operates within ethical boundaries.

Chile's neuro-rights legislation serves as a pioneering example, proving that it is possible to enshrine cognitive protections within constitutional and legal systems. Brazil's Proposta de Emenda Constitucional nº 29/2023, which seeks to incorporate the protection of mental integrity and algorithmic transparency into its constitutional framework, highlights the growing recognition of these issues in diverse legal contexts.

However, the challenges posed by neurotechnology extend beyond the legal domain. The commodification of neural data forces us to reevaluate fundamental concepts such as autonomy, dignity, and freedom. What does it mean to be human in an age where cognitive states can be measured, influenced, and sold? The mind, as the final frontier of privacy, must remain a sanctuary—a space free from intrusion and commodification. Upholding the sanctity of the human mind requires a collective commitment to ethical principles that prioritize individual dignity over economic gain.

The future of cognitive freedom depends on the actions taken today by policymakers, businesses, and civil society. Governments must enact robust laws to protect workers' mental privacy and cognitive autonomy, establishing clear boundaries for the ethical use of neurotechnologies. Businesses must adopt transparent, ethical practices that prioritize the well-being of employees over short-term productivity gains. Civil society must advocate for the recognition and enforcement of neuro-rights, ensuring that public discourse includes the voices of all stakeholders.

Neurotechnology offers humanity a glimpse into a future where cognition can be optimized, enhanced, and better understood. But this future must be guided by a steadfast commitment to the inviolability of the human mind. As societies grapple with the implications of these advancements, the question remains: Are we prepared to defend the freedom to think, dream, and define ourselves in a world where even our thoughts risk becoming commodities? The path forward demands vigilance, collaboration, and a shared resolve to prioritize humanity over exploitation. By embracing transparency, accountability, and respect for fundamental rights, we can ensure that neurotechnology becomes a force for progress that enhances human dignity rather than undermining it.

References

Baylis, F. (2011). *I Am Who I Am*: On the Perceived Threats to Personal Identity from Deep Brain Stimulation. Neuroethics, 6(3), 513-526.

Choudhury, S., & Slaby, J. (2011). *Critical Neuroscience*: A Handbook of the Social and Cultural Contexts of Neuroscience. Wiley & Sons, Incorporated, John.

Farahany, N. (2023). *Cognitive Liberty and the Future of Human Rights*. American Journal of Bioethics, 29(2), 110-125.

European Commission. (2022). *Declaration of Digital Rights and Principles for the Digital Decade*. COM(2022) 28 final.

Kant, I. (1964). *Groundwork of the Metaphysic of Morals*. New York: Harper & Row.

Lavazza, A. and Inglese, S. (2015). Free Will and Neurotechnology. *Journal of Ethics & Neuroscience*, 18(4), 201-217.

Martinez, W., Benerradi, J., Midha, S., Maior, H. A., and Wilson, M. L. (2022). Understanding the Ethical Concerns for Neurotechnology in the Future of Work. *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work* (CHIWORK '22), Article 17, 1–19.

OECD. (2021). *Neurotechnology and Society*: Ethical Considerations. OECD Publishing.

Ravinetti, R. (2023). *Legal Innovations*. Comparative Law Review, 29(2), 102-118.

Santos, A. L., & Costa, R. (2021). Neural Surveillance and Employee Rights. *International Journal of Labor Law*, 34(4), 128-145.

UNESCO. (2022). Ethics of Neurotechnology in the Workplace. UNESCO Digital Library.

UNESCO. (2023). *Neurotechnology and Human Rights*: Global Challenges. UNESCO Digital Library.

Yuste, R., et al. (2017). The Rise of Neuro-Rights. *Frontiers in Human Neuroscience,* 10(5), 87-102.

Zeki, S., et al. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London*. Series B: Biological Sciences, 359(1451), 1775-1785.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*: The Fight for a Human Future at the New Frontier of Power. New York, USA: Public Aff.

— 19 —

The future of neurotechnology: co-evolution between neuroethics and neurotechnology[1]

TAMAMI FUKUSHI
Tokyo Online University (Japan)
fukushi.tamami@internet.ac.jp

Introduction

In the past, human brain research was conducted in clinical settings by medical professionals involved in the diagnosis and treatment of patients with some kind of neurological or psychiatric condition. However, the situation surrounding human brain research has changed drastically since the end of the 20th century. In particular, the development and popularization of functional magnetic resonance imaging (fMRI), invented by Dr. Seiji Ogawa et al, has expanded the scope of research to healthy subjects, and research at facilities other than medical institutions has flourished. In addition, the spread of near-infrared spectrometers (NIRS) and improvements in electroencephalography (EEG) have reduced the physical burden of participation in experiments, and the age range of research subjects has expanded from infants to the elderly. Thus, human brain research has expanded from clinical medicine to cognitive psychology, and then to the elucidation of the neural basis of the human brain, including advanced thinking ability, morality, and ethics. This trend in neuroscience research has prompted the establishment of the academic field of Neuroethics. More than twenty years have passed since its establishment, and the information on brain activity obtained through the advancement of human brain measurement technology has been combined with the development of data science and artificial intelligence research to establish the technological field of neurotechnology.

[1] This work was supported by JSPS KAKENHI (Grant Number JP23K22293 and JP24K06393) and JST Moonshot JPMJMS2012.

Along with the progress of neurotechnology, neuroethics has also met informatics and law and is about to enter a new academic stage. This paper reviews the history of neuroethics, describes the development of neurotechnology and discusses the future ethical concerns it raises.

1. Synonyms for "Neuroethics" on the eve of its establishment

Neuroethics is a part of applied ethics that examines the ethical, social, and legal issues involved in the conduct of brain science research and the ethical use of research results. As an academic field, it was established in 2002 at the Dana Foundation conference called "Neuroethics: Mapping the Field" held in San Francisco. and the term rapidly spread thereafter (Marcus, 2002). Initially, Neuroethics was defined as "the ethics of neuroscience," which deals with practical ethical issues that arise in the process of conducting brain science research mainly on human subjects, and "the ethics of human society," which refers to the process of social value judgment and the ethical standards of human beings themselves, as a result of research findings on the neural basis of mental activities such as human-specific ethical views, religious beliefs, and value creation (Roskies, 2002). As a matter of fact, the arguments that led to the origin of these fields, and the analogy of neuroethics that reflects these arguments, were introduced more than 30 years before the establishment. The first appearance of the term "Neuro-ethics" was proposed by Pontius (Pontius, 1973). This was a critical review of an experimental paper by Zelazo and colleagues (Zelazo et al., 1972) that tested the hypothesis that neonates in the process of acquiring the gait reflex could acquire walking function earlier if they were given motor stimulation that would forcibly stimulate the gait reflex. The experimental approach "early intervention in the neurodevelopmental process that controls the motor system" is thought to have been used to argue that the conduct and results of the experiment were unethical in that they could interfere with the natural process of development of the entire nervous system and should be criticized (Fukushi and Sakura, 2009). The second synonym was "Neuroethicist" introduced by Cranford in 1989 (Cranford, 1989). This term appeared in a discussion of what role neurologists should play in light of their increasingly frequent involvement in ethical decisions in clinical medicine and institutional ethics committees, including patient care and end-of-life care. The third synonym, "Neuroethical," refers to the proximity of philosophy of mind and neuroscience and discusses the importance of neurological functions as the basis of human identity and the role of advanced neuroscience research in reexamining questions about the human psyche in the field of philosophical ethics (Churchland, 1991). The emergence of these terms was metaphorical, applied to the problems and their solutions in the extension of traditional bioethics and philosophy of science. Therefore, it cannot be said that they directly contributed to the creation and naming of Neuroethics as a modern academic field, but it is possible to seek the roots of the academic elements that Neuroethics encompasses (Illes, 2003).

2. Co-evolution of Neuroethics and Neurotechnology

Neuroscience has made many contacts with engineering fields, influenced by the entry of information and communication engineering, innovations in human brain

measurement technology, and the development of data science research using measurement data. In particular, neurotechnology, represented by technologies that measure brain activity and convert it into commands to operate equipment or linguistic information to be spoken, has made significant progress in recent years as a technology field that is expected to have market value due to its fusion with big data and AI technologies, as well as the lightweight and minimally invasive nature of devices used to measure brain activity. Therefore, from the perspective of Neuroethics, the practice of "responsible research and innovation (RRI)" that contributes to the ethical development, use, and diffusion of neurotechnology is progressed (OECD, 2019). Development of international standards for neurotechnology has also become a topic of discussion in the 2020s, with the Institute of Electrical and Electronics Engineers (IEEE) beginning to develop international standards for ethical aspects of technology development and dissemination (IEEE Standards Association, 2020, also see IEEE P7700 website https://standards.ieee.org/ieee/7700/11038/). The IEEE has also established an organization called the Neuroethics Framework, which considers the ethical use and dissemination of neurotechnology, taking into account the cultural aspects of the country or region to which the user belongs, depending on the purpose of use (https://brain.ieee.org/publications/ieee-neuroethics-framework/). Another organization, International Organization for Standardization/ International Electrotechnical Commission Joint Technical Committee1 (ISO/IEC JTC1) has also established Sub Committee SC43 to standardize Brain-Computer Interface (BCI) for non-medical purposes (see ISO/IEC JTC1 SC43 website https://www.iso.org/ committee/9082407.html). In this way, neuroethics related to governance in line with the development of neurotechnology, has co-evolved with various academic fields that encompass elements related to neuroethics, such as bioethics, philosophy of science, engineering ethics, law, religious studies, science and technology studies, and regulatory science, emphasizing the need to promote new systematization of neuroethics. Furthermore, as symbolized by the concept of Neurorights, it has become a driving force for the development of the fusion field of neuroethics and Neurolaw by increasing the points of contact between the fields of neuroethics and law (Chandler, 2018; Ienca 2021). Based on these circumstances, the next section discusses the relationship between the information environment and humans in the future that is predicted from the advancement of neurotechnology.

3. Future Possible Communication Technology Utilized by Advanced Neurotechnology

It has been pointed out that the development of Neurotechnology may bring about a major change in the forms of communication in human society. Research in archaeology, linguistics, psychology, biology, and other fields has revealed that humans have long communicated using verbal expressions by means of speech, symbols, letters, or gestures, and that these expressions are related to the biological evolutionary process followed by humans and the expansion of the cerebral cortex (Bretas et al., 2020). Neuroscience research has also revealed that there is a neural basis for human language ability and that the cerebral cortex contains brain sites responsible for language function (Friederici, 2011; Sousa et al., 2017). In recent years, the application of the BCI to the neural basis of language function has led to the development

of technologies for "verbalization of brain activity," i.e., reading internal language directly from the brain. This approach has led to research and development of brain-to-brain communication (B-to-B communication) technology (Nam et al., 2021; Silva et al., 2024). Such challenge offers new possibilities for speech synthesis research as an alternative means of communication for people with impaired speech functions due to nervous system diseases like amyotrophic lateral sclerosis (ALS), stroke, etc. The following is an overview of recent achievements of neurotechnological developments that may affect the changing forms of communication in human society and a discussion of possible future changes that may occur.

3.1. Speech synthesis

The attempts to "verbalization of brain activity" are closely related to the development of data science and language-generating artificial intelligence (AI) research, which have evolved dramatically over the past decade or so. A major contribution to this has been made by the neural decoding technique developed by Kamitani and his group regarding reproduction of visual information using MRI signals. This technique analyzes MRI signals in a wide range of brain regions by narrowly separated voxels and reconstructs what kind of information was received as a whole. They succeeded not only in reproducing information actually seen, from simple symbols and figures to photographs of people and landscapes, but also in reproducing visual information of arbitrary objects that were imagined (Kamitani and Tong, 2005; Horikawa and Kamitani, 2017). Similar research and development have been done with auditory information. Studies have been conducted on the reproduction of general auditory information, such as human speech and horse neighs (Santoro et al., 2017) and music (Bellier et al., 2023), as well as on the reproduction of information from the activity of neurons to read aloud specific sentences and then convert this information into artificial speech (Anumanchipallalli et al, 2019). Written language has also been studied in recent years, with "hands-free writing", in which the alphabet is estimated from the brain activity of a subject recalling a writing action and written out on a computer screen (Willet et al., 2021) and with estimating the meaning of sentences from brain activity in subjects who read silently a series of sentences presented visually (Tang et al., 2023). In addition, a technology that analyzes the neural activity of a subject recalling the "answer" by artificial speech to a question by listening to audio information of a question has made (Moses et al., 2019), and the one that analyzes the brain activity of a stroke victim while recalling what he/she wants to say and not only converts it into artificial speech but also links the mouth movements in the facial image to create a computer image as if the patient were speaking (Metzger et al., 2023) were reported .

3.2. Brain-Computer Interface utilizing Avatar Control

BCI is "the procedures and engineering elements involved in detecting, recording, analyzing, and converting nervous system activity into meaningful information that can be used to control external devices and computer elements " (Fukushi, 2024), and its research and development have continued for half a century since its concept was proposed in 1973 (Vidal, 1973; Wolpaw et al., 2021). BCI has been expanded to manipulate computer cursors (Hochberg et al., 2006), robotic arms (Hochberg et al., 2012),

avatars on computer screens (Luu et al., 2017) and humanoids (Chamora et al., 2020). The aforementioned "hands-free writing" (Willet et al., 2021) and artificial speech linking mouth movements (Metzger et a., 2023) are also part of BCI-operated avatars. The Cybernetic Avatar (CA) is a research and development of avatar operation technology, which is being conducted to improve the performance of avatars, to freely control avatars that reflect human intentions in real time in physical and cyber space (virtual reality space), to give social and economic value to their behavior (Hagita et al., 2024), and BCI is expected to contribute as a kind of control method for the CA.

3.3. Brain-to-Brain Communication

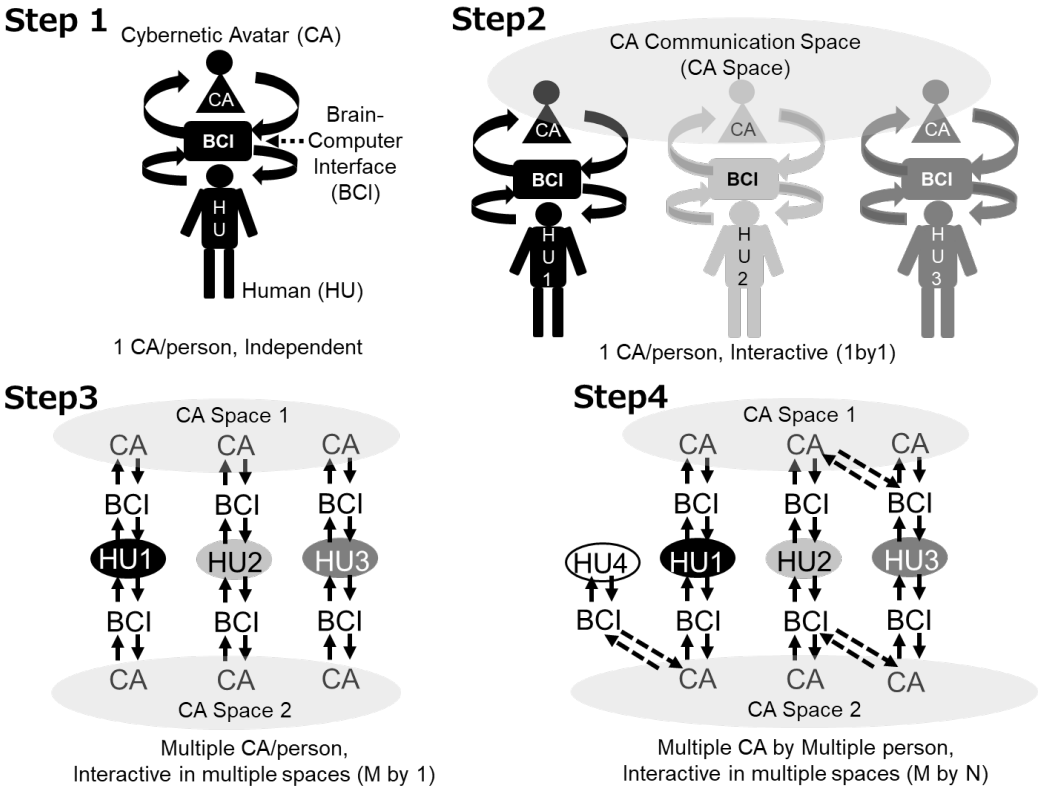
B-to-B Communication is a procedure of directly extracting brain information from one's brain and transmitting it to the other without involving the peripheral nervous system (Vakilipour and Fekrvand, 2024), and its development began in earnest in the 2010s, and after successful communication between animals of the same species (Pais-Vieira et al., 2013) and between humans and animals (Yoo et al., 2013), applications to humans only began in earnest (Vakilipour and Fekrvand, 2024; Nam et al., 2021). In 2014, a successful human-to-human B-to-B Communication experiment was reported in which brain information from one subject detected by EEG was transmitted to the other subject by transcranial magnetic stimulation (TMS)-induced expression of phosphenes (Grau et al., 2014). Furthermore, B-to-B communication between humans has reached a level where it is not just about conveying information, but where the person receiving the information is able to carry out some kind of behavioral task based on the information sent by the person sending it. For example, this includes behavioral tasks such as the movement of the right hand to "click" on a touch pad (Rao et al., 2014) and the "20 questions" task, in which the participants are divided into questioners and answerers and must obtain the correct answers (Stocco et al., 2015). This means that, in addition to the BCI function for transmitting brain information externally, they have also succeeded in mediating the Computer-Brain Interface (CBI) function for returning information received from others to the brain. Such facts indicate that B-to-B Communication is not just about sending and receiving information but has the potential to be a means of communication for multiple people to cooperate and accomplish a purposeful mission based on brain information. On this basis, the potential for future military applications, such as communication tool between multiple soldiers, has also been pointed out (Pei and Schalk, 2012; Maksimenko et al., 2018).

3.4. Future Communication Style with Advanced Neurotechnology

Advanced linguistic communication tools based on neurotechnology indicate the potential to greatly enhance the feasibility of B-to-B communication in human society. As I have described so far, speech synthesis, which extracts brain activities related to inner speech, is now possible, and furthermore, it is theoretically possible to express them via BCI-controlled CA, or to transmit them directly to the other party's brain through B-to-B Communication without expressing them as linguistic expressions. Considering these circumstances, one can examine the transformative potential of the forms of communication in human society in the future by providing step-by-step examples in this section.

Figure 1 schematically illustrates the expected phased transformation of communication through CAs. The first stage of communication is a relatively simple form in which an individual manipulates a CA through BCI to reflect his/her intentions and linguistic information on the CA and transmit them to others in the real world (Step 1 in Figure 1). This is already a rudimentary technology that has been achieved. The second stage is a form that involves communication between multiple CAs operated by individuals in either a virtual or real space and a space for communication specific to CA (CA Communication Space) is established (Step 2 in Figure 1). In the third stage, a single person operates multiple CAs and participates in communication in different spaces (Step 3 in Figure 1). In the fourth stage, it is expected to be transformed into an M by N form of communication, in which multiple people operate a certain CA(s) one time (Step 4 in Figure 1).

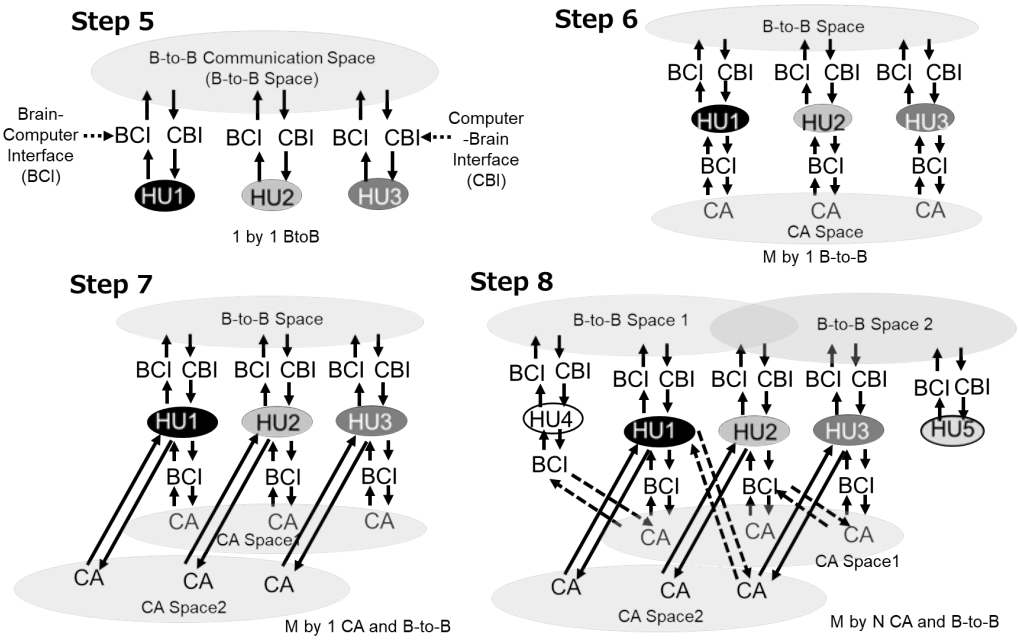
Figure 1
Examples of step-by-step evolution of communication style using CA controlled by human brain via BCI



The transformation does not stop here, as the application of B-to-B Communication is expected to make the situation more sophisticated and complex. Figure 2 is a schematic diagram of this process. In the fifth stage, one person participates in a single B-to-B Communication venue (this venue will inevitably be a virtual space) as shown in Step 5 in Figure 2. And in the sixth stage, one person participates in multiple spaces through B-to-B communication (Step 6 in Figure 2). What happens beyond that? As the seventh stage, a form of communication in which a single person participates in one communication space by BCI-controlled CA and the other by B-to-B Communication in one time is anticipated (Step 7 in Figure 2), which is still

relatively simple. In the eighth stage, one person could participate in multiple communication spaces by operating multiple CAs in addition to participate in multiple B-to-B communication spaces at one time. This might be close to the final and ultimate form of the transformation of communication in human society brought about by neurotechnology that is currently conceivable.

Figure 2
Examples of step-by-step evolution of communication style using CA controlled by human brain via BCI and B-to-B Communication



4. Ethical Concerns for Advanced Neurotechnologies

So far, I have discussed the possibility of transforming the forms of communication in human society through neurotechnology from a technological perspective. These transformations and their fundamental technologies, such as speech synthesis and B-to-B communication, have social significance in compensating for and overcoming communication disorders and securing the rights of people with disabilities (United Nations General Assembly, 2006). On the other hand, it has been pointed out that various ethical concerns exist. One of the most prominent concerns is the violation of "unreachability of the mind" or "freedom of thought" by extracting and inferring internal language from brain activity. It represents the risk that thoughts and feelings that should normally remain within the individuals will be expressed and communicated to others without permission, and there are also concerns that this may be a violation of "cognitive liberty". (Rainey et al., 2020; Yildirim-Vranckaert, 2024). As a solution to these concerns, the idea that information and events derived from the activities of one's neural system should be added to the list of protected fundamental human rights has been proposed, with some researchers proposing a separate rights title, "Neurorights" (Ienca, 2021). In addition to the terms introduced above, many other terms (e.g., Mental Integrity, Mental Privacy) have already been used to describe respect for

freedom and protection of rights related to the human mind, thought, and cognition. This situation has led to a discussion to clarify the interpretation of various terms and their relationship to neurorights, as well as to debate whether the neurorights should be recognized as an independent right and its effectiveness. The concept of neurorights or the right to mental and to ideology differs from country to country and from culture to culture, and the need to promote international discussion while respecting this diversity has also been pointed out (Herrera-Ferrá et al., 2023).

Chandler's group pointed that more practical implications of the widespread application and adoption of speech synthesis include " Testimony: Communication as Evidence," " Consent and Capacity: Communication as Means of Agency and Participation " and " Communication as Harm " (Chandler et al., 2022), which mean "the impact on communication as testimony and evidence in court," "the impact on autonomous decision-making and consent procedures (in clinical practice and research participation)," and "the risk of (and responsibility for) harm to oneself and others (due to expressed linguistic information)". These points are coupled with the technical concern that it may not be possible to distinguish whether the linguistic information originated from speech synthesis was truly intended to be verbalized by the individual or kept internal (Rainey et al, 2019). At the moment, as speech synthesis by neurotechnology reported so far relies on brain activities recorded from the regions associated with speech movements (Anumanchipalli et al., 2019; Metzger et al., 2023), one can consider that only linguistic information with speech intentions would be considered to be represented (Chandler et al., 2022). However, there is no certain technical evidence for this interpretation, and it does not address the concerns associated with the expansion of brain regions targeted for research and development and the integration of technologies such as B-to-B Communication that do not necessarily require information transfer between brain regions related to speech movements. In the future, the reliability and ethics of the process of converting brain information into linguistic information will be important factors. Careful language generation technology development and algorithm design based on shared ethical and moral values between the individual providing the brain information and the provider of the technology or service that converts it into linguistic information will be required (Maslen and Rainey, 2019; van Stuijvenberg et al., 2024).

The fact that the development of speech synthesis has a significant impact on the development of B-to-B Communication indicates that there are many common ethical concerns between the two, but there are also ethical concerns that will become more serious as B-to-B Communication develops and expands. As shown in Figure 2, while CA control requires BCI only, B-to-B Communication requires two interfaces, BCI and CBI, to be connected to an individual's brain. In addition, in the case of communication from in which CA control and B-to-B communication are mixed, as in Step 6, or in the case of communication by participating in multiple CA spaces and B-to-B spaces, as in Steps 7 and 8, to what extent can the human brain withstand the simultaneous processing of multiple information environments? In other words, how much can the human brain handle the simultaneous processing of multiple information environments or can the "I" personality in each communication space retain its self-identity, or does it split into different selves that depend on the information environment in that communication (Lyreskog et al., 2024). This would be a question that depends on the standards and algorithms of the artificial intelligence in the BCI and CBI used to pro-

cess the respective information, and at the same time, on the relationships with others participating in the communication space(s) and their changes. As mentioned above, given the possibility that B-to-B Communication could be used for military purposes, there might be a need to reconsider the pros and cons of neurotechnology contributing to the enforcement of combat actions and the performance of missions that do not meet one's wishes, depending on the difference in the position of the communication partner and one's own position, and the command structure in the team, as well as reconsidering the scope of responsibility of individual soldiers and the professional ethics based on this (Latheef, 2023). The transformation of communication style brought about by advances in neurotechnology raises philosophical questions about the dignity of human mental processes and the identity of the human personality, as well as legal and ethical issues such as the protection of human rights regarding mental freedom, the admissibility of evidence, and the assignment of responsibility. Neurotechnology and neuroethics need to co-evolve, with an awareness of the progress and maturation of ethical debates in step with the advancement and diffusion of technology.

Conclusion

The characteristics of the elemental technologies encompassed by neurotechnology and the applications expected to be implemented in society are diverse, and the ethical concerns surrounding these applications and the laws and regulations to be considered must also be diverse. In order to deal with these issues smoothly and effectively, it is important for all stakeholders involved in neurotechnology to acquire appropriate literacy and to work on improving their literacy in accordance with technological developments and the emergence of ethical issues. The development of advanced science and technology sometimes leads people to imagine the future according to outlandish ideas or overly aggressive bad scenarios, and there might be a risk of falling into ethical disputes for which no realistic solutions can be found. In line with the life cycle of products and services that are expected to apply neurotechnology, country bodies and international organizations would be required to take strategic measures to improve the ability to acquire detailed literacy according to the position and role of stakeholders, such as researchers, developers, providers, users, those who play a mediating role, regulatory authorities, and administrative authorities. In other words, strategic measures for capacity building would be required by national governments and international organizations. It is essential that whole stakeholders have a proper understanding of the development of neurotechnology, and that accurate assessment of technological progress with calm discussions would be proceeded while incorporating moral and philosophical perspectives.

References

Anumanchipalli, G.K., Chartier, J., Chang, E.F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753), 493-498. DOI: https://doi.org/10.1038/s41586-019-1119-1

Bellier, L., Llorens, A., Marciano, D., Gunduz, A., Schalk, G., Brunner, P., Knight, R.T. (2023). Music can be reconstructed from human auditory cortex activity using nonlinear decoding models. *PLoS Biolology*, 21(8), e3002176. DOI: https://doi.org/10.1371/journal.pbio.3002176

Bretas, R.V., Yamazaki, Y., Iriki, A. (2020). Phase transitions of brain evolution that produced human language and beyond. *Neuroscience Research*, 161, 1-7. DOI: https://doi.org/10.1016/j.neures.2019.11.010.

Bublitz, J.C. (2022). Novel Neurorights: From Nonsense to Substance. *Neuroethics*, 15(1), 7. DOI: https://link.springer.com/article/10.1007/s00440-002-0236-0

Chamola, V., Vineet, A., Nayyar, A., Hossain, E. (2020). Brain-Computer Interface-Based Humanoid Control: *A Review*. Sensors (Basel), 20(13), 3620. DOI: https://doi.org/10.3390/s20133620

Chandler, J.A. (2018). Neurolaw and Neuroethics. *Cambridge Quarterly of Healthcare Ethics*, 27(4), 590-598. DOI: https://doi.org/10.1017/S0963180118000117

Chandler, J,A,, Van der Loos, K.I., Boehnke, S., Beaudry, J.S., Buchman, D.Z., Illes, J. (2022). Brain Computer Interfaces and Communication Disabilities: Ethical, Legal, and Social Aspects of Decoding Speech From the Brain. *Frontiers in Human Neuroscience,* 6, 841035. DOI: https://doi.org/10.3389/fnhum.2022.841035

Churchland, P.S. (1991). Our Brains, Our Selves. Reflections on Neuroethical Questions. In: Roy, D.J., Wynne, B.E., Old, R.W. (eds) *Bioscience-Society*. New York: John-Wiley and Sons, p. 77–96.

Cranford, R.E. (1989). The neurologist at ethics consultant and as a member of the institutional ethics committee. The neuroethicist. *Neurologic Clinics,* 7(4), 697–713.

Díaz Soto, J,M,, Borbón, D. (2022). Neurorights vs. neuroprediction and lie detection: The imperative limits to criminal law. *Frontiers in Psychology*, 13, 1030439. DOI: https://doi.org/10.3389/fpsyg.2022.1030439

Friederici, A.D. (2011). The brain basis of language processsing: from structure to function. *Physiological Reviews*, 91(4), 1357-1392. DOI: https://doi.org/10.1152/physrev.00006.2011

Fukushi, T. (2024). East Asian Perspective of Responsible Research and Innovation in Neurotechnology. https://www.sciencedirect.com/science/article/pii/S2667242124000435. DOI: https://doi.org/10.1016/j.ibneur.2024.04.009

Fukushi, T., Sakura, O. (2008). Exploring the Origin of Neuroethics: From the Viewpoints of Expression and Concepts. *American Journal of Bioethics,* 8(1), 56-57. DOI: https://doi.org/10.1080/15265160701839672

Grau, C., Ginhoux, R., Riera, A., Nguyen, T.L., Chauvat, H., Berg, M., Amengual, J.L., Pascual-Leone. A., Ruffini, G. (2014). Conscious brain-to-brain communication in humans using non-invasive technologies. *PLoS One*, 9(8), 105225. DOI: https://doi.org/10.1371/journal.pone.0105225

Hagita, N., Kanai, R., Ishiguro, H., Minamizawa, K., Arai, F., Shimpo, F., Matsumura, T., Yamanishi, Y. (2024). Cybernetic avatars: Teleoperation technologies from in-body monitoring to social interaction. *Sci Robot*. 9(96), eadg1842. DOI: https://www.science.org/doi/10.1126/scirobotics.adg1842

Herrera-Ferrá, K., Muñoz, J.M., Nicolini, H., Saruwatari Zavala, G., Martínez Bullé Goyri, V.M. (2023). Contextual and Cultural Perspectives on Neurorights: Reflections Toward an International Consensus. *AJOB Neuroscience*, 14(4), 360-368. DOI: https://doi.org/10.1080/21507740.2022.2048722

Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., Donoghue, J.P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099), 164–171. DOI: https://doi.org/10.1038/nature04970

Hochberg, L.R., Bacher, D., Jarosiewicz, B., Masse, N.Y., Simeral, J.D., Vogel, J., Haddadin, S., Liu, J., Cash, S.S., van der Smagt, P., Donoghue, J.P. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398), 372-375. DOI: https://doi.org/10.1038/nature11076

Horikawa, T., Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, 8, 15037. DOI: https://doi.org/10.1038/ncomms15037

IEEE Standards Association (2020). STANDARDS ROADMAP: Neurotechnologies for Brain-Machine Interfacing. https://standards.ieee.org/wp-content/uploads/import/documents/presentations/ieee-neurotech-for-bmi-standards-roadmap.pdf

Ienca, M. (2021). On Neurorights. *Frontiers in Human Neuroscience*, 15, 701258. DOI: https://doi.org/10.3389/fnhum.2021.701258

Illes, J. (2003). Neuroethics in a new era of neuroimaging. *American Journal of Neuroradiology*, 24(9), 1739-1741. https://www.ajnr.org/content/24/9/1739

Kamitani, Y., Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5), 679-685. DOI: https://doi.org/10.1038/nn1444

Latheef, S. (2023). Brain to Brain Interfaces (BBIs) in future military operations; blurring the boundaries of individual responsibility. *Monash Bioethics Review*, 41(1), 49-66. DOI: https://doi.org/10.1007/s40592-022-00171-7

Ligthart, S., Ienca, M., Meynen, G., Molnar-Gabor, F., Andorno, R., Bublitz, C., Catley, P., Claydon, L., Douglas, T., Farahany, N., Fins, J.J., Goering, S., Haselager, P., Jotterand, F., Lavazza, A., McCay, A., Wajnerman, Paz. A., Rainey, S., Ryberg, J., Kellmeyer, P. (2023). Minding Rights: Mapping Ethical and Legal Foundations of 'Neurorights'. *Cambridge Quarterly of Healthcare Ethics*, 15, 1-21, DOI: https://doi.org/10.1017/S0963180123000245

Luu, T.P., Nakagome, S., He, Y., Contreras-Vidal, J.L. (2017). Real-time EEG-based brain-computer interface to a virtual avatar enhances cortical involvement in human treadmill walking. *Scientific Reports*, 7(1), 8895. DOI: https://doi.org/10.1038/s41598-017-09187-0 s

Lyreskog, D.M., Zohny, H., Mann, S.P., Singh, I., Savulescu, J. (2024). Decentralising the Self - Ethical Considerations in Utilizing Decentralised Web Technology for Direct Brain Interfaces. *Science and Engineering Ethics*, 30(4), 28. DOI: https://link.springer.com/article/10.1007/s11948-024-00492-2

Maksimenko, V.A., Hramov, A.E., Frolov, N.S., Lüttjohann, A., Nedaivozov, V.O., Grubov, V.V., Runnova, A.E., Makarov, V.V., Kurths, J., Pisarchik, A.N. (2018). Increasing Human Performance by Sharing Cognitive Load Using Brain-to-Brain Interface. *Frontiers in Neuroscience*, 12, 949. DOI: https://doi.org/10.3389/fnins.2018.00949

Maslen, H., Rainey, S. (2021). Control and Ownership of Neuroprosthetic Speech. *Philosophy & Technology*, 34(3), 425-445. DOI: https://doi.org/10.1007/s13347-019-00389-0

Marcus, S.J. (ed.) (2002). *Neuroethics*: Mapping the Field. Dana Press, Washington, DC.

Metzger, S.L., Littlejohn, K.T., Silva, A.B., Moses, D.A., Seaton, M.P., Wang, R., Dougherty, M.E., Liu, J.R., Wu, P., Berger, M.A., Zhuravleva, I., Tu-Chan, A., Ganguly, K., Anumanchipalli, G.K., Chang, E.F. (2023). A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, 620(7976), 1037-1046. DOI: https://doi.org/10.1038/s41586-023-06443-4

Moses, D.A., Leonard, M.K., Makin, J.G., Chang, E.F. (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*, 10(1), 3096. DOI: https://doi.org/10.1038/s41467-019-10994-4

Nam, C.S., Traylor, Z., Chen, M., Jian,g X., Feng, W., Chhatbar, P.Y. (2021). Direct Communication Between Brains: A Systematic PRISMA Review of Brain-To-Brain Interface. *Frontiers in Neurorobotics*, 15, 656943. DOI: https://doi.org/10.3389/fnbot.2021.656943

OECD (2019). Recommendation of the Council on Responsible Innovation in Neurotechnology. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457

Pais-Vieira, M., Lebedev, M., Kunicki, C., Wang, J., Nicolelis, M.A. A brain-to-brain interface for real-time sharing of sensorimotor information. *Scientific Reports*, 3(1319), DOI: https://doi.org/10.1038/srep01319

Pei, X., Hill, J., Schalk, G. (2012). Silent communication: toward using brain signals. I*EEE Pulse*, 3(1), 43-6. DOI: https://doi.org/10.1109/MPUL.2011.2175637

Pontius, A.A. (1973). Neuro-ethics of "walking" in the newborn. *Perceptual and Motor Skills,* 37(1), 235–245. DOI: https://doi.org/10.2466/pms.1973.37.1.235

Rainey, S., Martin, S., Christen, A., Mégevand, P., Fourneret, E. (2020). Brain Recording, Mind-Reading, and Neurotechnology: Ethical Issues from Consumer Devices to Brain-Based Speech Decoding. *Science and Engineering Ethics*, 26(4), 2295-2311. DOI: https://doi.org/10.1007/s11948-020-00218-0

Rainey, S., Maslen, H., Mégevand, P., Arnal, L.H., Fourneret, E., Yvert, B. (2019). Neuroprosthetic Speech: The Ethical Significance of Accuracy, Control and Pragmatics. *Cambridge Quarterly of Healthcare Ethics*, 28(4):657-670. https://doi.org/10.1017/S0963180119000604

Rao, R.P., Stocco, A., Bryan, M., Sarma, D., Youngquist, T.M., Wu, J., Prat, C.S. (2014). A direct brain-to-brain interface in humans. *PLoS One,* 9(11), e111332. DOI: https://doi.org/10.1371/journal.pone.0111332

Roskies A. (2002). Neuroethics for the new millennium. *Neuron*, 35(1), 21-23. DOI: https://www.cell.com/fulltext/S0896-6273(02)00763-8

Santoro, R., Moerel, M., De Martino, F., Valente, G., Ugurbil, K., Yacoub, E., Formisano, E. (2017). Reconstructing the spectro-temporal modulations of real-life sounds from fMRI response patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 114(18), 4799-4804. DOI: https://doi.org/10.1073/pnas.1617622114

Sousa, A.M.M., Meyer, K.A., Santpere, G., Gulden, F.O., Sestan, N. (2017). Evolution of the Human Nervous System Function, Structure, and Development. *Cell,* 170(2), 226-247. DOI: https://doi.org/10.1016/j.cell.2017.06.036

Tang, J., LeBel, A., Jain, S., Huth, A.G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 26(5), 858-866. DOI: https://doi.org/10.1038/s41593-023-01304-9

United Nations General Assembly, (2006). *United Nations General Assembly Session* 61 Resolution 106, Convention on the Rights of Persons with Disabilities, A/RES/61/106. New York, NY: United Nations General Assembly, http://un-documents.net/a61r106.htm

Vakilipour, P., Fekrvand, S. (2024). Brain-to-brain interface technology: A brief history, current state, and future goals. *International Journal of Developmental Neuroscience*, 84(5), 351-367. DOI: https://doi.org/10.1002/jdn.10334

van Stuijvenberg, O.C., Samlal, D.P.S., Vansteensel, M.J., Broekman, M.L.D., Jongsma, K.R. (2024). The ethical significance of user-control in AI-driven speech-BCIs: a narrative review. *Frontiers in Human Neuroscience,* 18, 1420334. DOI: https://doi.org/10.3389/fnhum.2024.1420334

Vidal, J.J. (1973). Toward direct brain-computer communication. *Annual review of biophysics and bioengineering*, 2, 157-180. DOI: https://doi.org/10.1146/annurev.bb.02.060173.001105

Willett, F.R., Avansino, D.T., Hochberg, L.R., Henderson, J.M., Shenoy, K.V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, 593(7858), 249-254. DOI: https://doi.org/10.1038/s41586-021-03506-2

Wolpaw, J.R., Millán, J.R., Ramsey, N.F. (2020). Brain-computer interfaces: Definitions and principles. *Handbook of Clinical Neurology*, 168, 15-23. DOI: https://doi.org/10.1016/B978-0-444-63934-9.00002-0

Yildirim-Vranckaert, E.O. (2024). The Right to Construct Yourself and Your Identity: The Current Human Rights Law Framework Falls Short in Practice in the Face of Illegitimate Interference to the Mind. *American Journal of Law & Medicine*, 49(2-3), 267-285. DOI: https://doi.org/10.1017/amj.2023.31

Yoo, S.S., Kim, H., Filandrianos, E., Taghados, S.J., Park, S. (2013). Non-invasive brain-to-brain interface (BBI): establishing functional links between two brains. *PLoS One*, 8(4), e60410. DOI: https://doi.org/10.1371/journal.pone.0060410

Zelazo, P.R., Zelazo, N.A., Kolb, S. (1972). "Walking" in the newborn. *Science*, 176(4032), 314-315. DOI: https://doi.org/10.1126/science.176.4032.314

— EPILOGUE —

— 20 —

Neurolaw in action: toward to the nearest future of Law and Neuroscience

KEIGO KOMAMURA
Keio University, Tokyo
komamura@keio.jp

In the past few years, neuroscience and its technologies have attracted attention as commodity with the trade name like "neurotech" or "braintech" that has spread a variety of benefits and concerns throughout our society. In response, legal and ethical issues have been identified, and new laws, regulations, and ethical standards are urgently needed.

Various papers in this volume were written with such global trends in mind. They raise new ideas, agendas, and challenges, and they refer to various modes of legal response.

This article closes the book with a brief perspective on this new "neurolaw" movement.

1. ELSI on the Neuroscience and its Technology

I have long been a ordinary constitutional law scholar with very traditional research style. However, I became involved in the Moonshot Research & Development Project[1] by JST (the Japan Science and Technology Agency) starting in 2021. The project seeks to innovate both theory and technology in neuroscience and brings together Japan's leading natural scientists specializing in neuroscience, engineering, mathematics, and medicine. There, various neurotechnologies are being proposed, developed, and implemented in society to improve our wellbeing in the near future. For example, BMI (brain-machine interface) has been successfully applied to the treatment of ALS and spinal cord injury patients to restore their lost motor functions

[1] Moonshot Research & Development Project (https://www.jst.go.jp/moonshot/en/index.html) (visited on Feb. 10, 2025).

by activating compensatory circuits through neurotech, and is already being used in rehabilitation sites. In addition, our team planned and organized "Brain Pick," in which EEG is used to control an avatar on a monitor to play a virtual game. It is like Olympics where people compete using brain information. Thus, we are committed to entertainment using neurotech.

On the other hand, beyond wellbeing enhancement and entertainment, there are also activities that are more candid in their scientific ambitions and interests. The project which I am a part of calls itself the "Internet of Brains (IoB)".[2] As the name suggests, this group is ultimately aiming for a world in which brains are connected to brains through technology. In a sense, it is very ambitious and extraordinary project indeed.

A few years ago, I was asked to join to do so-called ELSI activities in this project, IoB, which has a wide scope ranging from the development of neurotech that solves specific problems that disturb our wellbeing to the realization of a somewhat SciFi-like vision of transformation of the human beings and the world.

As such, I have been leading ELSI research in neuroscience technology for four years. In this paper, I would like to make an epilogue to this book by reviewing the four years of progress and by writing down what I think about ELSI research in neurotech.

I would like to propose the following two topics. First, I would like to comment on the ELSI activities themselves. Second, to briefly comment on the recent international framework formation.

2. The Research Activity called ELSI

(1) Problems with the ELSI methodology

First, ELSI is too inclusive. As is well known, ELSI (ethical, legal, and social implications)[3] is an academic activity to evaluate and control science and technology using knowledge and discipline of humanity and social science such as ethics, law, and social sciences. In recent times, the "E" has been cubed including two more Es as economics and environmental studies, and the concept has apparently developed to include not only ethics but also economics and environmental protection.

But how can we confront science and technology with such a complicated perspective where many different and distinct disciplines compete? To begin with, each of the humanistic disciplines listed here has its own unique discipline, and it is very difficult to integrate them into a single viewpoint.

Despite this, people seem to think that those different disciplines in the human-social science fields can be easily integrated. It seems that there is an easy way of thinking that academic knowledge can be integrated as soon as it is put on the table together. In short, ELSI is (probably) not going very well because of the rampant bias that discards the individuality of each of the humanities and social sciences. Before we can go beyond difference between natural science and socio-humanities, we must

[2] Internet of Brains (https://brains.link/en) (visited on Feb. 10, 2025).

[3] About a definition and a brief history of ELSI research, see the website of the National Human Genome Research Institute (https://www.genome.gov/10001754/elsi-planning-and-evaluation-history) (visited on Feb.10, 2025).

integrate difference of disciplines in socio-humanities, which are lumped together under the name of ELSI.

In order to convert this into an effective perspective on the ELSI, I suggest that we abandon the conventional easy-going idea of "just gather all the information in one place," and narrow the research focus to the following three points, and conduct them separately.

(2) **A Proposal: Three Pillars for a new ELSI**

"Effective Design of Normative Control":

First, the ELSI should focus on "normative control" for neurotech, since ethics and law are both *normative* disciplines different from other fields of socio-humanities. One more thing to be added is "scientific norms. This scientific norm is a condition for science to be established itself as science, and in short, it demands "trust" in science. In other words, it is an activity to clarify whether the "efficacy" and "benefit" of a neurotech is really experimentally proven. Specifically, it will be important to verify whether the research supporting neurotech is supported by appropriate evidence.

"Promoting Social Utility":

The normative approach is not the only role that social science should play. Activities that increase individual or social productivity, reduce social costs, or increase social utility through neurotech are a social issue separate from normative control. I suppose those activities to promote social utility are more rational to conduct on their own.

"Search for Social Acceptance":

The third point is the search for social acceptance. Outreach activities that accurately and attractively convey the true state of the development of technology to society are another important aspect of ELSI. It will become increasingly important for the ELSI to create a unit of activities that can generate acceptance of neurotech by controlling both excessive anxiety and excessive expectations of new technologies.

(3) **From ELSI to RRI, and Outreach**

Since the above missions are different from each other, and the academic knowledge to be mobilized is also different, it is reasonable to conduct them separately. In addition, because the term ELSI is rather misleading and myriad, the above three pillars should be established under the concept of RRI (responsible research and innovation), which has been proposed to replace ELSI.

However, there may be aspects in which it is appropriate to integrate these three pillars and work together. One such aspect is how to incorporate international trends to make a global framework of neuroethics and neurolaw into domestic movement to make national laws and regulatory frameworks, which will be discussed later, and this will be a great opportunity for neurolaw. The three pillars mentioned above will work together to encourage domestic administrative and legislative efforts, which will be particularly compelling in cases where funding is sought. We are entering an era in which public outreach activities seeking legislation and funding are in strong demand.

3. Recent Trends in International Frameworks

In the past few years, various international organizations have proposed frameworks for regulating and promoting neurotech, including OECD, UNESCO, UNICEF, IEEE, and the UN Human Rights Council, etc. Many documents, including reports and recommendations in final/tentative draft form, have been published. I had the opportunity to comment on the UN Human Rights Council reports[4] at the Hybrid Minds conference in Geneva in 2024.[5] In what follows, I would like to list some points that should be noted for the future of neurolaw, mainly with the UN Human Rights Council's report.

First. Interspersed throughout the reports and recommendations are various concepts of "neurorights" as the basis for normative control. They are, of course, freedom of thought, mental privacy, personal integrity, and cognitive liberty, etc. These concepts of rights listed in those reports and recommendations have good reason to be taken up. Many of them already have a substantive basis in the Universal Declaration of Human Rights and the International Covenants on Human Rights, and in that sense, they are a globally shared list of rights.

However, the scope of these rights is by no means clear or stably applied. The UN Human Rights Council is trying to position these rights concepts as *actionable* rights, i.e., legally binding normative concepts that can be applied to the legal cases in court, but it is quite difficult to do so at present. In view of the fact that these international efforts will be transposed into national laws in the future, it is necessary to make further efforts to clarify the inherent core, scope, and degree of legally binding force of these unclear and controversial rights concepts.

Second. Each international document is slightly different in its envisioned normative control. Some focus on controlling and fostering neurotech businesses in the marketplace and the neurotech industry on a national scale, others are concerned with military applications and military use, and still others are concerned with educational and workplace applications. This trend is welcome from the perspective of carefully tailoring normative control to the characteristics of the subject. Also discussed at UNESCO was the widening disparities and strengthening of divisions brought about by *technological colonization*.[6] It is interesting to note that they tried to view this issue in the context of a kind of cultural imperialism. but this concept, like other concepts of rights, is contentious.

Third. Related to the point mentioned above, the *right to science* may be important. This is a rights concept that appears in only one place in the UN Human Rights Council report,[7] and is a part of international human rights law that has its basis in the Universal Declaration of Human Rights and the International Covenants on Human

[4] "Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights: report of the Human Rights Council Advisory Committee" (file:///C:/Users/81804/Downloads/A_HRC_57_61-EN.pdf)(visited on Feb.10, 2025).

[5] See, the website of Hybrid Mind Conference, Geneva, 16-18 October 2024 (https://memento.epfl.ch/event/hybrid-mind-conference-geneva-16-18-october-2024-2/) (visited on Feb.10, 2025).

[6] See, page 9 of "First draft of a Recommendation on the Ethics of Neurotechnology (revised version): working document as of 27 August 2024" (file:///C:/Users/81804/Downloads/391074eng.pdf) (visited on Feb.10, 2025).

[7] See, page 14 of "Impact, opportunities and challenges of neurotechnology with regard to the promotion and protection of all human rights: report of the Human Rights Council Advisory Committee" (supra note 4).

Rights. The impact that this right can have on science and technology is significant, including public access to information, disclosure requests, public participation, ensuring access for use, and fairness in the allocation of resources. One of the speakers at Hybrid Minds symposium emphasized the importance of this right to science, and I agree with that very much. However, this is also controversial because of its high impact and broader scope.

* * *

As the development of neurotech accelerates, the project to envision legal and ethical controls over it has just begun. However, as is clear from what has been mentioned so far, even at this point in time, many proposals have already been made, and various concepts of rights and regulatory frameworks are dancing on the stage with unclear and dark settings, without a common understanding.

The new field of jurisprudence, neurolaw, despite its many ambiguities and conflicts, is an attractive frontier before us.