COMMON HUMAN RIGHTS CHALLENGES RAISED BY DIFFERENT APPLICATIONS OF NEUROTECHNOLOGIES IN THE BIOMEDICAL FIELDS

Report commissioned by the
Committee on Bioethics (DH-BIO)
of the Council of Europe

Author: Marcello Ienca

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

COMMON HUMAN RIGHTS CHALLENGES RAISED BY DIFFERENT APPLICATIONS OF NEUROTECHNOLOGIES IN THE BIOMEDICAL FIELD

All requests concerning the reproduction or translation of all or part of this document should be addressed to the Directorate of Communication (F-67075 Strasbourg Cedex).

All other correspondence concerning this document should be addressed to the Directorate General of Human Rights and Rule of Law.

Layout: SPDP, Council of Europe
© Council of Europe, October 2021

ABOUT THE AUTHOR

Dr. Marcello Ienca is a Principal Investigator at the College of Humanities at EPFL where he leads the ERA-NET funded Intelligent Systems Ethics research unit. He is also an affiliate member of the Health Ethics and Policy unit, Department of Health Sciences and Technology, and an ordinary member of the Competence for Rehabilitation Engineering & Science at ETH Zurich, Switzerland.

Dr. Ienca's scholarship focuses on the ethical, legal, social and policy implications of emerging technologies. In particular, he investigates the broader implications of new (and often converging) sociotechnical trends such as Artificial Intelligence (AI), big data, digital epidemiology, robotics, assisted living, digital health, social media, dual use, and neurotechnology. He and his team use both theoretical and empirical methods to explore the requirements for responsible innovation, ethically aligned technology design, user-centred design, and human-centered technology assessment.

Dr. Ienca is actively involved in science and technology policy within international organizations and professional societies. In particular, he is a member of the Organisation for Economic Co-operation and Development's (OECD) Steering Group on Neurotechnology and the representative of the Swiss Delegation (appointed by the Swiss State Secretariat for Education, Research and Innovation, SERI). He has also been invited to serve as an expert advisor to the Council of Europe's Ad Hoc Committee on Artificial Intelligence and the Bioethics Committee. Dr. Ienca has written reports for the OECD, the Council of Europe, and the European Parliament's Panel for the Future of Science and Technology. He is a Member of the Board of Directors of the Italian Neuroethics Society (SINe), a former Board Member and current member of the Nominating Committee of the International Neuroethics Society (INS).

Ienca is a member of the Editorial Board of several academic journals such as Neuroethics, Bioethica Forum, Frontiers in Neuroergonomics and Frontiers in Genetics. Ienca has received several awards for social responsibility in science and technology such as the Vontobel Award for Ageing Research (Switzerland), the Prize Pato de Carvalho (Portugal), the Sonia Lupien Award (Canada), the Paul Schotsmans Prize from the European Association of Centres of Medical Ethics (EACME).

He has authored one monograph, several edited volumes, more than 60 scientific articles in peer-review journals, several book chapters and is a frequent contributor to Scientific American. His research was featured in academic journals such as Neuron, Nature Biotechnology, Nature Machine Intelligence, Nature Medicine and media outlets such as Nature, The New Yorker, The Guardian, The Times, Die Welt, The Independent, the Financial Times and others.

ACKNOWLEDGEMENTS

The author would like to thank all the people who either participated in the preliminary expert consultation, commented on earlier drafts of this report, or contributed to shaping the ideas herein contained, in particular: Tom Douglas, Nita Farahany, Fabrice Jotterand, Philipp Kellmeyer, Fruzsina Molnar Gabor, Jennifer Chandler, Jan-Christoph Bublitz and Rafael Yuste. Furthermore, the author would like to thank the Council of Europe DH-BIO Secretariat, in particular, Laurence Lwoff and Natalia Zaytseva, for the fundamental coordination work without which this report would have not been possible.

This report has been commissioned and supported by the Committee on Bioethics of the Council of Europe. The report is based on a combination of original work and previous scholarship of the author on the human rights implications of neurotechnology. See:

- Ienca, M. (2021). On neurorights. *Frontiers in Human Neuroscience, 15: 701258*.
- Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life sciences, society and policy*, *13*(1), 1-27.
- Ienca, M. (2017). The Right to Cognitive Liberty. *Scientific American*, *317*(2), 10-10.
- Ienca, M. (2019). *Intelligenza2: per un'unione di intelligenza naturale e artificiale*. Rosenberg & Sellier.

TABLE OF CONTENTS

GLOSSARY

- **Alpha waves:** neural oscillations in the frequency range of 8-12 Hz.
- **Belief:** An attitude that some proposition about the world is true.
- **Biomarker**: a biological marker, i.e., a measurable indicator of some biological state or condition.
- **Brain (human)**: The central organ of the human nervous system.
- **Brain function:** the function of neuronal circuits in the brain.
- **Cognition:** the set of mental processes such as thinking, knowing, remembering, judging, and problem-solving.
- **Cognitive enhancement:** Interventions in the brain that improve cognition (e.g., attention, concentration, and information processing in executive functions such as reasoning and decision-making).
- **Cognitive liberty:** the right of individuals to make free and competent decisions regarding their minds and brains.
- **Locked-in syndrome**: A neurological condition in which the patient is alert and conscious but is unable to move or communicate verbally due to complete paralysis of almost all voluntary muscles of the body with the exception of vertical eye movements and blinking.
- **Mental content:** the content of a mental state, either conceptual or non-conceptual.
- **Mental integrity:** the integrity of the human mind.
- **Mental privacy**: people's right against the unconsented intrusion by third parties into their brain data as well as against the unauthorized collection of those data.
- **Neurorights**: Ethical, legal, social or natural principles of freedom or entitlement related to a person's cerebral and mental domain.
- **Neurodiscrimination:** discrimination based on neural features.
- **Neuroimaging:** the use of various techniques to either directly or indirectly image the structure, function, or pharmacology of the nervous system.
- **Neurostimulation:** purposeful modulation of the nervous system's activity using invasive (e.g., microelectrodes) or non-invasive means (e.g. transcranial magnetic stimulation or transcranial electric stimulation).
- **Neurotechnology**: the broad and heterogenous spectrum of methods, systems and instruments that establish a connection pathway to the human brain through which neuronal activity can be recorded and/or altered.
- **Personhood**: the status of being a person as opposed to a nonperson.
- **Psychological continuity**: people's continuity of their mental life over time (e.g., continuity across non-synchronous mental states).

EXECUTIVE SUMMARY

Neurotechnologies are emerging technologies that establish a connection pathway to the human brain through which human neuronal activity can be recorded and/or altered. These technologies open novel opportunities for exploring, influencing, or intercommunicating with the human brain. Medical neurotechnologies offer the potential to help people with neurological or psychiatric conditions such as Parkinson's disease, dementia, stroke, and major depressive disorder. Non-medical neurotechnology systems provide new tools and methods to monitor and modulate brain activity in healthy subjects and to interact with digital devices. Intervening effectively and safely in the human brain through neurotechnology is a scientific frontier that must be reached for the good of humanity. At the same time, however, it raises major ethical and legal challenges. Neuroethics and neurolaw are the two main areas of scholarship that address, respectively, the ethical and legal issues raised by our ever-improving ability to intervene in the brain through neurotechnology.

In the past decade, philosophical-legal studies in the fields of neuroethics and neurolaw have given increasing prominence to a normative analysis of the ethical-legal challenges in the mind and brain sciences in terms of rights, freedoms, entitlements, and associated obligations. This way of analyzing the ethical and legal implications of neuroscience has come to be known as "neurorights". Neurorights can be defined as the ethical, legal, social, or natural principles of freedom or entitlement related to a person's cerebral and mental domain; that is, the fundamental normative rules for the protection and preservation of the human brain and mind. In their most popular version, neurorights have been defined as an emerging category of human rights designed to protect the brain-mind sphere of the person.

Reflections on neurorights have received ample coverage in the mainstream media and have become a mainstream topic in the public neuroethics discourse. Further, they are rapidly becoming an emerging regulatory tool of international politics. Yet, several meta-ethical, normative-ethical, legal-philosophical and practical challenges need to be solved to ensure that neurorights can be used as effective instruments of global neurotechnology governance and be adequately imported into international human rights law. To overcome these challenges, this report attempts to provide a comprehensive normative-ethical, historical and conceptual analysis of neurorights. In particular, the objective of this report is fivefold as it attempts to (i) provide an overview of current and likely future biomedical neurotechnologies; (ii) reconstruct a history of neurorights and situate these rights in the broader history of ideas; (iii) summarize ongoing policy initiatives related to neurorights in the present international policy landscape; (iv) proactively address some unresolved ethical-legal challenges; and (v) identify priority areas for further academic reflection and policy work in this domain.

The findings of this report suggest that neurorights reflect fundamental human interests that are deeply rooted in the history of ideas. These rights introduce normative specifications related to the protection of the person's cerebral and mental domain that are not merely repetitive of existing human rights frameworks, but add a new, fundamental level of normative protection. This corroborates the view that human beings generally enjoy a set of rights against certain kinds of interferences in their brains and minds, including those interferences involved in the misuse of neurotechnologies. In addition to protecting against the misuse of neurotechnology, the neurorights spectrum also contains moral and legal provisions aimed at ensuring that neuroscientific and neurotechnological progress is used to empower people and improve human well-being (positive rights). To a large extent, the findings of this report also corroborate the normatively stronger thesis that the fundamental rights and freedoms relating to the human brain and mind should be seen as the fundamental substrate of all other rights and freedoms.

This overview indicates that there is not yet complete consensus regarding the conceptual-normative boundaries and terminology of neurorights. Divergences exist in relation to how these rights are interpreted, named, and conceptually articulated. Nonetheless, some degree of convergence is emerging around three main families of neurorights. First and foremost, the need for specific provisions on the protection of private brain-related information seems to share a high degree of acceptance and recognition. The right to mental privacy appears to be the candidate best equipped conceptually to take on this role. Second, the right to mental integrity appears to have the highest degree of legal entrenchment. While there are some variations in the interpretation of this right, there is full theoretical consensus about the need to protect the person from psychological harm and mental interference. Third, a variety of neurorights candidates have been proposed to preserve and promote the freedom of the human mind and thereby prevent external manipulation. These include evolutionary interpretations of the right to freedom of thought, the right to cognitive liberty, and the right to personal identity.

On the other side of the coin, positive rights such as promoting justice and equality—e.g., through ensuring egalitarian access to neurotechnology for biomedical use and promoting patient welfare on the basis of the ethical principle of beneficence—have so far occupied a secondary role in the neurorights debate.

Introducing neurorights into the human rights framework may require adding new protocols to existing instruments or even stipulating new multilateral instruments entirely devoted to neuroethics and neurolaw. In either case, some fundamental ethical, meta-ethical, and legal issues must be addressed in order to overcome problems such as rights inflation and to provide an adequate normative justification for neurorights. These include introducing justificatory tests for the introduction of neurorights, clarifying the relationship between moral and legal neurorights and harmonizing neurorights with existing normative instruments.

The Council of Europe’s Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (Oviedo Convention) offers an ideal platform and normative substrate for the protection and promotion of neurorights. Given its focus on prohibiting the misuse of innovations in biomedicine, protecting the dignity and identity of all human beings, and guaranteeing respect for their integrity and fundamental freedoms, the Convention is well placed for either enshrining neurorights through ad hoc protocols or for serving as a basis for future instruments.

Understanding, treating, and augmenting the human brain and mind is one of the great scientific challenges of our age. Achieving these goals in a way that preserves justice, safeguards fundamental rights and human dignity is the corresponding task of ethics and law. Neurorights will likely be a useful tool to accomplish this task.

1. INTRODUCTION: THE NEUROTECHNOLOGY REVOLUTION

Neurotechnology is the umbrella term commonly used to describe a broad and heterogenous spectrum of methods, systems and instruments that establish a connection pathway to the human brain through which neuronal activity can be recorded and/or altered. In other words, we can define neurotechnology any technology for exploring, influencing or intercommunicating with the human brain.

Over the past three decades, technological innovation and scientific discoveries in the fields of neuroscience and biomedical engineering, combined with concomitant progress in computer modelling and machine learning software for data analysis, have led to cumulative progress in neurotechnology. This progress has resulted in a broad spectrum of clinical applications for the prevention, diagnosis, treatment, and care of people with neurological and psychiatric disorders. Due to the pace of such progress, authors have called this “the neurotechnology revolution” (Scott, 2013). Furthermore, the reduction in hardware costs and the increase in the size of the neurotechnology market, have recently favoured the spillover of neurotechnological applications into various extra-clinical areas of human activity. These include the judicial, military, communication, and consumer industry sectors. This extraclinical diffusion has been accompanied by increased reflection, debate, and, more recently, deliberation about the ethical, legal, and social implications of neurotechnology. Due to both the technological advancement and social diffusion of neurotechnology applications, public interest in neurotechnology has increased in recent years.

This report has a fivefold objective. First, it takes stock of current and likely future neurotechnology systems and methods. Second, it provides a critical overview of the current and foreseeable ethical and legal implications of neurotechnology, with particular focus on their human rights implications and the emerging category of “neurorights”. Third, it summarizes ongoing policy initiatives related to neurorights in the international policy landscape. Fourth, it proactively addresses some unresolved ethical-legal challenges related to the introduction of neurorights into the human rights framework. Finally, it identifies priority areas for further academic reflection and policy work in this domain.

The first section of this report provides a summary of the history of neurotechnology, an overview of current neurotechnological systems and methods, and a critical appraisal of the capabilities and limitations of current neurotechnologies.

1.1. A HISTORY AND TAXONOMY OF NEUROTECHNOLOGY

1.1.1. Neuroimaging

The history of modern neurotechnology begins at the end of the 19th century. Since the second half of the nineteenth century, various techniques have been developed to detect components of brain activity. In 1873, the Italian physiologist Camillo Golgi discovered a method of chromatographic impregnation for visualising neurons under the microscope. Two years later, the English physiologist Richard Canton used an electromechanical instrument called a galvanometer (named after the Bolognese anatomist Luigi Galvani) to observe electrical impulses from the surfaces of the living brains of a rabbit and a monkey. Forty-nine years later, in 1924, the German psychiatrist Hans Berger recorded the first human electroencephalography (EEG). With this instrument of his own invention, Berger was able to observe and measure electrical activity in the human brain and discover alpha waves, i.e., neural oscillations in the frequency range of 8-12 Hz, which later proved to be very important in neurofeedback processes. Since then, the EEG has been widely used in clinical and research settings as an electrophysiological monitoring method for recording the electrical activity of the brain. This electrical activity is recorded non-invasively, i.e., from outside the skull: electrodes are placed along the scalp to measure voltage fluctuations resulting from the ionic current within the neurons of the brain.

Not only electricity but also magnetism has helped open up the frontiers of brain activity to scientific observation. In 1971, US chemist Paul Lauterbur applied magnetic field gradients in all three dimensions and a back-projection technique to create images of physical objects. The first images obtained using this technique were published in 1973 in the journal *Nature* marking the first time that magnetic fields and radio waves were being used to create images of physical bodies (Lauterbur, 1973). Lauterbur called his method 'zeugmatography imaging', a term that was later replaced by the more common **'magnetic resonance imaging' (MRI)**. MRI soon proved to be very useful for visualising internal, and therefore not easily visible, body structures. Clearly, the human brain was one preferred subject. The first ever MRI scan of the human brain was captured, in 1982, by the US scientist John Schenck using a magnet rated at 1.5 tesla, hence strong enough to penetrate the human body and obtain clear, high-resolution images.

Brain function refers to the function of neuronal circuits in the brain. It includes the neural correlates of mental states such as thoughts, experiences and actions: given numerous repetitions of a thought, experience or action, it is possible to use statistical methods to reliably determine which areas of the brain had undergone a change in magnetic signal, i.e. which areas of the brain were most active during the realization of that thought, action or experience.

Figure 1- Brain function

The depiction of anatomical maps of the brain, however, is insufficient to observe brain *mechanisms*, that is its processes and functions. To make such observations, new neurotechnologies had to be developed. Thanks to the pioneering research of the physiologist Angelo Mosso it was already hypothesized, at the end of the 19th century, that local blood flow within the brain was intimately linked to brain function. Mosso had observed that when neurological patients were engaged in cognitive tasks such as mathematical calculations, brain pulsations increased locally. Towards the end of the 1980s, researchers at the Massachusetts General Hospital confirmed and documented this increase in local cerebral blood flow in the areas of greatest neuronal activity through high-resolution MRI. The method they used exploited images of the primary visual cortex by subtracting cerebral blood volumes calculated in the stimulated states and comparing them with the unstimulated states, in order to obtain functional maps showing increased activity in the areas of the brain subjected to stimulation. Almost at the same time, Seiji Ogawa of the University of Tokyo in Japan had shown that it was possible to detect the different magnetic properties of haemoglobin in its oxygenated and deoxygenated forms using magnetic resonance imaging. By exploiting these properties, it became possible to observe variations in the magnetic signal within the brain. With this new method, known today as Blood Oxygen Level-Dependent Imaging (BOLD), a new technique called **functional magnetic resonance imaging (fMRI)** was derived from MRI. fMRI allows to measure brain activity indirectly, i.e., using haemodynamic (cerebral blood flow) responses as indirect markers. Current fMRI techniques can localise brain activity, visualise brain activation patterns graphically and determine their intensity by colour-coding the strength of activation. These techniques are now used for a variety of purposes including pre-operative risk assessment and functional mapping of brain areas to detect functional abnormalities (e.g., left-right hemispheric asymmetry in language and memory regions) or to observe patients' post-stroke or post-operative recovery. fMRI has also been used to monitor the effects of drug and behavioural therapies on patients, or even to aid the diagnosis of neurological conditions such as depression and Alzheimer's disease (de Vos et al., 2018; Li, Xu, & Lu, 2018).

Since the birth of fMRI, a neurotechnological revolution has taken place. Over the last thirty years, the number, variety, and degree of precision of technological tools capable of visualising, recording or even modifying brain activity has been increasing at an ever-increasing rate. During the 1990s, often referred to as the “decade of the brain”, the use of neuroimaging techniques increased dramatically. In addition to EEG, MRI and fMRI, a number of other techniques were added over time, including positron emission tomography (PET), magnetoencephalography (MEG), and functional near-infrared spectroscopy (fNIRS). PET, now one of the main techniques of nuclear medicine, is used to observe metabolic processes in the body, brain included. It detects pairs of gamma rays emitted indirectly by a radioligand (a radioactive biochemical substance) capable of emitting positrons (the anti-particles of electrons), which is introduced into the body through a biologically active molecule called a 'radioactive tracer'. The application of this technique to the study of the brain is based on the assumption that areas of increased radioactivity are associated with increased brain activity. As with fMRI, what is indirectly measured in PET is the flow of blood to different parts of the brain, which is generally thought to correlate with neuronal activity. PET is now used for a variety of diagnostic purposes including the search for brain tumours, the diagnosis of Alzheimer's disease or pre-surgical preparation for epilepsy.

MEG, in contrast, allows to map brain activity by recording the magnetic fields produced by naturally occurring electrical currents in the brain using highly sensitive magnetometers. First measured by physicist David Cohen in 1968, MEG was originally carried out in a magnetically shielded room to reduce magnetic background noise. Today's MEG sensors (about three hundred) are placed inside a helmet, which covers most of the head of the person. In this way, MEG measurements can now be collected more quickly and efficiently. More recently, prototypes of wearable MEG headsets have been developed, i.e., MEG systems that can be worn like a helmet, allowing free and natural movement during scanning instead of requiring the patient to remains still within a restrictive scanner (Boto et al., 2018).

fNIRS was discovered by Frans Jobsis in 1977 based on the observation that biological tissues are relatively transparent to light in the near-infrared region, i.e. between 700 and 1300 nanometres, making it possible to transmit enough photons through body organs to monitor these organs in situ (Jobsis, 1977). In this near-infrared region, haemoglobin—in its two main variants: oxyhaemoglobin (O2Hb) and deoxyhaemoglobin (HHb)—exhibits an oxygen-dependent absorption mode. It therefore acts as a chromophore (a substance responsible for absorbing electromagnetic radiation in the visible range) in biological tissue that absorbs light in this near-infrared region. Since it does not require particularly bulky or heavy hardware, fNIRS is now used in a wide variety of portable devices and applications that involve monitoring human brain activity in relation to behavioral performance in natural environments and everyday conditions.

Neurotechnologies such as EEG, fNIRIS (and possibly, in the near future, also portable MEG and fMRI) are transforming neuroimaging technology by making it suitable for widespread use. This is resulting in the development of miniaturized, battery-operated, wireless sensors that can measure brain activity in ambulatory and field settings (Ayaz, Izzetoglu, Izzetoglu, & Onaral, 2019).

Finally, over the last 10 years, **optogenetics** has become increasingly used in neuroscience «for the study of how specific cell types contribute to brain functions and brain disorder states” (Boyden, 2015). This technique involves the use of light to control neurons that have been genetically modified to express light-sensitive ion channels. Although optogenetics toolsets are not fully mature and currently have limited application in the clinic, they can be increasingly used as neural connectivity and cell phenotyping tools as well as neural recording and imaging tools. It has been argued, however, that «the results emerging from the use of optogenetics in basic neuroscience, and from neurotechnology as a whole, will provide in the years to come a variety of insights into new molecular targets for drug development, new circuit sites for electrical brain stimulation, new protocols of regenerative medicine, and other strategies for helping repair the brain» (Boyden, 2015).

1.1.2. Intracranial electrophysiological monitoring

Unlike all previous techniques, intracranial electrophysiological monitoring does not measure brain activity from the outside but from inside the skull. The most common form of intracranial electrophysiological monitoring is electrocorticography (ECoG), also called intracranial electroencephalography (iEEG). Although this technique uses electrodes to record electrical potentials (similar to electroencephalography), these electrodes are not placed on the scalp but directly on the exposed surface of the brain, that is on the cortex itself. Since it requires a neurosurgical procedure, ECoG is unsuitable for studying the brains of healthy subjects—where the collateral risks would be too high compared to the potential benefits—but it is widely used in animal models and human brain surgery, as part of pre-surgical exploration to determine the location and limits of pathological foci, especially in the areas most prone to epileptic phenomena. In terms of measurement quality, being able to place electrodes directly on the cortex makes ECoG much more precise, informative and detailed than EEG and opens up a much larger window of access to human cognitive processes.

Since 2017, a private company named Neuralink started working on an intracranial electrophysiological monitoring system that reads information from the brain. The system leverages probes composed mostly of polyimide, with a thin gold or platinum conductor. The probes are to be inserted into the brain through an automated process performed by a surgical robot. Each probe consists of an area of wires that contains electrodes capable of locating electrical signals in the brain, and a sensory area where the wire interacts with an electronic system that allows amplification and acquisition of the brain signal. Each system makes of up to 3072 electrodes per formation. The technology has been tested so far in animal models (rodents, suidae and monkeys) but the company has anticipated starting experiments in humans shortly.

1.1.3. Brain-Computer Interfaces for motor control and communication

Neuroimaging techniques were originally developed and are still largely implemented in diagnostics in order to detect abnormalities in the brain of neurological patients such as lesions, tumours, or indicators of epileptic episodes. Since the end of the 1960s, however, a branch of neurotechnology has been developing with the aim of not only 'reading' brain activity but also exploiting it to create a direct communication channel with digital computers. This field of research is known as brain-computer interface (BCI) and its birth was partly due to copious investments by the US Department of Defence and extensive research with animal models. In 1969, a single neuron in the brain of a monkey was successfully connected to an external device placed in front of the monkey's face. The device was able to move according to the neuron's activity, causing the movement to be triggered whenever the monkey was actively thinking about something. After several training sessions, the animal had learnt to intentionally activate one of its neurons in order to trigger the activity of an external device. The official invention of brain-computer interfaces, however, is usually dated four years later, in 1973, when Jacques Vidal published a paper on brain-machine communication entitled 'Toward direct brain-computer communication' (Vidal & JJ, 1973). In this paper, the term 'brain-computer interface' or BCI was coined for the first time and a prototype of a direct communication channel between the brain and a computer was presented. The first demonstrations of Vidal's BCIs were based on the control of movements in a simulated maze. Fifteen years later, researchers at the University of Skopje succeeded for the first time in using a BCI to control a physical robot via brain activity.

In the 1970s and 1980s, various studies succeeded in establishing increasingly reliable communication channels between the brain activity of non-human animals and computer systems. Towards the end of the 1990s, the first invasive interfaces for humans with motor disabilities were developed. Researchers at Emory University in Atlanta, led by Philip Kennedy and Roy Bakay, were the first to install a brain implant in a human being that produced signals of sufficient quality to simulate the patient's desired movements. The patient, a 50-year-old American, suffered from what is known as 'locked-in' syndrome. The implant was installed in the patient's brain in 1998 and allowed him to learn to control a computer cursor through his own brain activity.

Locked-in syndrome:
A neurological condition in which the patient is alert and conscious but is unable to move or communicate verbally due to complete paralysis of almost all voluntary muscles of the body with the exception of vertical eye movements and blinking. The patient's consciousness is usually confirmed by his ability to understand verbal language and respond to interlocutors through his own eye movements, in fact the only window to the world of a locked-in patient. Moreover, the EEG of a locked-in patient is usually the same as that of a healthy patient.

Figure 2- Locked in syndrome

Over the subsequent two decades, several research groups have made progress in transforming data generated by brain electrodes into outputs of actions performed by a computer. Today's BCIs allow quadriplegic patients to use common tablets, write emails, chat, stream music or watch videos on YouTube solely through brain activity. Recently, a quadriplegic patient was able to play Beethoven's "Ode to Joy" by directly commanding a piano simulation app with his thoughts (Nuyujukian et al., 2018). In addition to cursors and tablets, digital devices that can now be controlled directly with the brain also include robotic arms (e.g., for patients who have had a limb amputated), electronic wheelchairs, or even entire humanoid robots.

Modern BCIs can be divided into two types: **invasive** and **non-invasive**. Invasive BCIs record brain signalling by surgically implanting electrode arrays inside the brain or are otherwise directly connected to the central nervous system. Non-invasive BCIs interface brain signalling with neuroimaging technologies such as electroencephalography (EEG) that record brain activity through electrodes placed outside the skull. Both invasive and non-invasive BCIs establish a direct interaction between the user's brain and a computing device. This interaction is usually described as a four-step cycle.

The first phase concerns **input**, i.e., the generation of specific brain activity by the user in response to a stimulus. This brain activity is generated when the user is in a certain cognitive state or performs a certain mental task. For example, when a BCI user controls an electronic wheelchair, a matrix of possible route choices is presented on the interface the user is looking at. A frequent brain activation pattern used in BCI are the so-called event-related potentials (ERPs), i.e., measured brain responses that are the direct result of a specific sensory, cognitive or motor event. Among these ERPs, a growing interest is surrounding the P300 wave, an ERP component usually elicited in decision making (Fazel-Rezai et al., 2012). In our example, when the desired route is presented at the interface, the user's brain signals will contain a P300 signal that can be detected by the BCI.

The second phase concerns the **measurement and recording of brain activity**. In this phase, brain activity patterns in the user's brain are detected and measured by the interface during a cognitive process or the performance of a mental task. For example, when a BCI user chooses a certain route option (e.g., a specific end position, or an instruction to turn left) the BCI can detect the P300 wave in real time. The recorded measurement can be implemented in different ways depending on the type of BCI in use. The most common type of non-invasive BCI is based on electroencephalography (EEG); other measurement options include functional near-infrared spectroscopy (fNIRS) or invasive solutions such as electrocorticography (ECoG).

In order to be usable for BCI and generate appropriate outputs (i.e., those desired by the user), the raw neural data measured in the second phase must be decoded into their main features and subsequently classified. **This decoding and classification** process typically takes place in the third phase of the BCI cycle. In this phase, the data are processed to 'clean up' the brain signals, i.e., to increase the signal-to-noise ratio (i.e., the measure of the strength of the desired signal relative to the background noise) and to filter out the most relevant aspects of each signal for further processing. This processing is necessary to extract the relevant features from the signal and distinguish them from irrelevant features, especially from the background noise due to the underlying brain activity that is not directed at performing that specific mental task. In this process of decoding and classification, artificial intelligence plays a fundamental role. Machine learning algorithms are generally trained to correctly decode the recorded and amplified neural signals, as well as to correctly classify the resulting data categories.

Once decoded, the signals can be translated into output. The output is usually the execution of the action initially intended or desired or deemed beneficial to the user through the control of the applications interfaced by the BCI (in our example, turning left in a wheelchair). Controllable applications include motorised devices (e.g., electronic wheelchairs and robotic limbs), sensor devices and various software and hardware applications (including smartphone applications). Once each cycle is completed, the user can perceive feedback from the previous cycle (e.g., experience that the wheelchair turns left) and the next cycle can begin.

1.1.4. Neurostimulation

Neurotechnology makes it possible not only to read brain activity, but also to *rewrite* neural signals. The brain not only emits electrical signals that can be recorded directly or indirectly by machines, but it is also capable of receiving electrical signals from inside or outside the brain. This type of neurotechnology that can modulate brain activity is usually called '**neurostimulation**' or ‘neuromodulation’ depending on the mode of operation. Neurostimulation works by applying electrodes to the brain, the spinal cord or peripheral nerve. Neuromodulation works by either actively stimulating nerves to produce a natural biological response or by applying targeted pharmaceutical agents in tiny doses directly to site of action. While the two terms are often used interchangeably, “neuromodulation” tends to emphasize a broader and slower change in brain function, such as in the context of disease treatment.
Neurostimulation and neuromodulation can be either invasive or non-invasive.

Of the invasive methods, the most common and most promising is **deep brain stimulation** (DBS). DBS is the main form of surgical treatment used to treat the motor symptoms of neurological disorders such as Parkinson's and dystonia. Recently, this tecnique also showed positive results in mitigating the symptomps of treatmentresistent major depressive disorder (https://www.nature.com/articles/s41591-021-01480-w). Being an invasive practice, it requires the implantation of thin wires with electrodes inside the brain. These electrodes are connected to extension leads that are channelled under the skin behind the patient's ear and along the neck. The extensions under the skin are in turn connected to a pulse generator (a device similar to a cardiac pacemaker), usually placed under the skin of the chest or above the stomach. A remote control allows the patient to turn the pulse generator on and off. When the device is switched on, the electrodes provide high-frequency stimulation of the affected area. This stimulation modifies some of the electrical signals in the brain that cause Parkinson's symptoms, especially tremor. Over the past two decades, neurotechnology research has shown that DBS can reduce symptoms and improve the quality of life for people with advanced Parkinson's whose symptoms are no longer manageable with medication. Although DBS is usually implanted in the late stages of Parkinson's, recent studies have found that earlier use of deep brain stimulation can have beneficial effects. In some cases, DBS is able to eliminate the tremor and allows the person in question to regain the motor functions necessary to carry out daily activities such as washing, eating, drinking, reading and so on. The last-generation DBS systems leverage intelligent computers interconnected with the brain that automatically adjust the level of stimulation according to each person's symptoms. This is called **closed-loop DBS.** In traditional open-loop DBS, a neurologist needs to manually adjust the stimulation parameters every 3-12 months after DBS implantation. On the other hand, in last-generation closed-loop DBS, the programming of the stimulation parameters is performed automatically by the device based on the measured biomarker.

Since electric current can pass through the skull, neurostimulation and neuromodulation can also be released non-invasively. This property is widely exploited by a broad spectrum of technologies called transcranial electrical stimulation. These include transcranial direct current stimulation (tDCS) and transcranial alternating current stimulation (tACS). Both tDCS and tACS were originally developed for the treatment of patients with brain injuries or psychiatric conditions such as depression. However, due to their non-invasive nature, both technologies have been tested for cognitive enhancement in healthy people. Over the past few years, some studies have found evidence of moderate but significant cognitive improvements associated with transcranial electrical stimulation. One of the most important of these studies was conducted by neuroscientists Robert Reinhart and John Nguyen of Boston University, who used a combination of EEG and tACS to monitor and then non-invasively stimulate the brains of two groups of participants. The first group was made up of people aged between 20 and 29. The second group included people much older, aged 60 to 76. By comparing the two groups' performance on work and memory tasks via EEG, the researchers found that the older individuals were slower and less accurate at remembering objects they had seen or identifying subtle differences between two almost identical images. The older group also exhibited less synchronisation of brain activity between two parts of the brain highly involved in memory processes: the frontal and temporal cortex. By stimulating both the frontal and temporal cortex for 25 minutes using tACS, the elderly subjects not only improved the synchronisation of brain activity, but also their performance on memory tests, to the point where their scores were comparable to those of the group of 20-year-olds. This effect lasted for over fifty minutes after stimulation before returning to normal levels. Using this type of stimulation, the researchers had not only demonstrated that they could reconnect or re-synchronise memory circuits, but also that they could temporarily turn some aspects of the biological clock of human memory back forty years (Reinhart & Nguyen, 2019). Results of this kind demonstrate the possibility, at least in principle, of using neurotechnology not only to decode brain activity or to control objects with thought, but even to enhance human cognitive abilities beyond their threshold of (statistically speaking) biological normality: a phenomenon generally described by the phrase **'cognitive enhancement'** (Bostrom & Sandberg, 2008).

1.1.5. Pervasive Neurotechnology

Neurotechnologies were originally developed and are still mostly implemented in the context of clinical medicine and neuroscience research. In recent years, however, a number of neurotechnology applications have made their way into the market and are now integrated into a number of consumer devices for healthy users for various non-clinical purposes. The umbrella term usually used to encapsulate all these non-invasive, scalable and potentially ubiquitous applications of neurotechnology is **'pervasive neurotechnology**' (Ienca, Haselager, & Emanuel, 2018), a notion borrowed from the more widely used notion of 'pervasive computing'. Today, pervasive neurotechnology applications include BCIs for device control or real-time neuromonitoring, neurosensor-based vehicle operator systems, cognitive training tools, electrical and magnetic stimulation of the brain, wearables for mental well-being, and virtual reality systems.

Most of these pervasive applications use EEG recordings or other non-invasive techniques to monitor electrical activity in the brain for a variety of purposes including neuromonitoring (real-time assessment of brain function), neurocognitive training (using certain frequency bands to improve neurocognitive function) and device control. EEG-based BCIs are increasingly used as wearable accessories for a range of everyday activities including gaming, entertainment, and remote smartphone control. For example, the companies Emotiv, Neurosky and Muse offer a wide range of wireless headsets for everyday use that can be connected to compatible smartphones and personal computers. Brain control can be used to remotely control different types of devices and engage in different activities including gaming and other forms of entertainment, marketing, self-monitoring, and communication. Thanks to pervasive neurotechnology, the application of brain-computer interfaces to device control is no longer limited to the clinical field. Recently, for example, the car manufacturer Nissan presented a kind of brain-car interface that detects a driver's brain waves in order to monitor his attention span, thus anticipating and preventing possible lapses in attention due to fatigue or sleep. The aim of the system is to allow the vehicle to respond a few fractions of a second faster than the driver's natural reaction time. According to the car manufacturer, this technology can improve not only safety but also driving quality even in experienced drivers. Another device called UDrone, manufactured by the Chinese company EEGSmart, allows to control a drone with brain activity. The UDrone is a small, lightweight device with 2-inch props, is equipped with an 8-megapixel camera and can stay in the air for about 7 minutes. Brain activity is read via EEG and translated into a wireless signal with which to control the device.

The possibility of non-invasive brain control has raised the attention also of the mobile communication industry. Several leading companies including Apple and Samsung are incorporating neurogadgets into the accessory assortments of their major products. For example, iPhone accessories such as the XWave headset already allow users to connect directly to compatible iPhones and read brain waves. Meanwhile, prototypes of the next generation of Samsung Galaxy Tabs and other mobile or wearable devices have been tested to be controlled by brain activity via EEG-based BCI (Powell, Munetomo, Schlueter, & Mizukoshi, 2013). In light of these trends, some experts predicted that neurodevices will gradually replace the keyboard, touch screen, mouse, and voice command device as humans' preferred ways to interact with computers.

Not only neuroimaging devices and BCIs fall into the category of pervasive neurotechnologies. Various electrical brain stimulators also fall into this category too. Unlike neuroimaging devices, neurostimulators are not primarily used to record or decode brain activity, but rather to stimulate or modulate brain activity electrically. Portable and easy-to-use transcranial direct current stimulation (tDCS) devices are the most popular form of consumer neurostimulator. They are used in a variety of low-cost direct-to-consumer applications aimed at optimising brain performance on a variety of cognitive tasks, depending on the brain region being stimulated. Recently, transcranial magnetic stimulation (TMS)—a magnetic method used to briefly stimulate small regions of the brain for both diagnostic and therapeutic purposes—has also evolved into portable devices, which have been found to be effective in the treatment of migraine (Lefaucheur et al., 2014).

Since 2018, tech giant Facebook has dedicated a team of 60 engineers to work on building a brain-computer interface that will allow users to write posts on the social network using only their minds, thus bypassing keyboard and voice commands. The team plans to use optical imaging techniques to scan the brain a hundred times a second to detect the silent language in a person's head and translate it into text. Regina Dugan, head of Facebook's Building 8 R&D division, explained that the goal is to enable people to type a hundred words a minute, five times faster than they would on a phone. This technology, dubbed 'direct brain interface' by Facebook, will combine machine learning algorithms for decoding natural language with advanced spatial resolution optical neuroimaging systems and next-generation neurodevices. However, the future of this technology is uncertain since Facebook has recently announced in a blog post that they are discontinuing the project and will instead focus on an experimental wrist controller for virtual reality that reads muscle signals in the arm. Other companies that operate in the pervasive neurotechnology domain include the aforementioned Neuralink and Kernel. The latter company has developed a variety of brain activity monitoring devices and software. Devices include “Flux” and “Flow”. Software products include "Sound ID," a program that can tell what speech or song a person is listening to just from brain data.

In short, while the last few decades have seen neurotechnology unlock the human brain and make it readable under a scientific lens and utilizable for medical purposes, the decade just begun will likely see neurotechnology become pervasive and integrated into numerous aspects of our lives and increasingly effective in modulating the neural correlates of our psychology and behaviour. While we should welcome continued progress in the development of neurotechnologies, the ethical and legal implications of the neurotechnology revolution should be considered early and proactively. As neurotechnologies advance, it is crucial to assess whether our current human rights framework is conceptually and normatively well equipped to address the new challenges arising in the brain-computer-society intersection, thus, to simultaneously provide guidance to researchers and developers, while providing protection to individuals and groups.

1.2. WHAT INFORMATION CAN NEUROTECHNOLOGY READ?

Although very different in functionality, applicability, and mode of utilization, the various neurotechnologies described above share at least one common characteristic: they can record quantitative data about human brain structure, activity and function. These include direct measurements of brain structure, activity and/or function (e.g., neuronal firing or summed bioelectric signals from EEG) and indirect functional indicators (i.e., blood flow in fMRI and fNIRS). These quantitative data about the structure, activity and function of the human brain can be called ‘**human brain data’**. Human brain data can reveal information about a person health status (e.g., neurological, or psychiatric health) and, to some extent, support inferences about mental processes.

Information about brain and mental health:

> A biomarker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (Biomarkers Definitions Working Group. 2001).
>
> *Figure 3- Definition of Biomarker*

As we have seen, neuroimaging techniques can be used to image the morphology (e.g., MRI), the function (e.g., fMRI), the metabolism (e.g., PET), or the molecular content (e.g., MR spectroscopy) of the human brain. The data generated through these neuroimaging techniques can be used to identify image-based biomarkers of brain disease, hence called **neuroimaging biomarkers**. For example, neuroimaging biomarkers of

Alzheimer's Disease (AD) include measurement of beta-amyloid deposition with amyloid PET or of brain and hippocampal atrophy with MRI (Bateman et al. 2012). Neuroimaging biomarkers can also be used to identify early signatures of cognitive decline (Gordon et al., 2018). Neuroimaging biomarkers of multiple sclerosis (MS) include counting and volume of lesions, enhancing lesions, and black holes. Neuroimaging biomarkers of ischemic stroke include, at the acute phase, the volume of the ischemic penumbra, as estimated by MR perfusion-based parameters. More recently, neuroimaging data have been used to also identify correlates of mental disorders such as depression and schizophrenia (Castanheira, Silva, Cheniaux, & Telles-Correia, 2019). Biomarkers of neurological, psychiatric and mental disorders are sensitive data, as they can go to the very core of a human being, their health and their mental life.

Information about mental states: At the neurobiological level, brain data are the most direct correlates of mental states. Current neurotechnologies, especially non-invasive techniques, are not yet able to *decode thoughts*. This means that current neurotechnologies are not capable of providing a full, granular and real-time account of the neural patterns of specific cognitive processes. However, they already allow to infer the engagement of some mental (e.g. perceptual and cognitive processes) from patterns of brain activation, through a process known as *reverse inference* (Poldrack, 2011). This occurs through invasive and non-invasive methods to record (and manipulate) neuronal circuits as well as ML-driven data analytics. In laboratory animals, proof-of-concept studies have shown the possibility of decoding visual perception and manipulating it with high precision (Carrillo-Reid, Han, Yang, Akrouh, & Yuste, 2019; Marshel et al., 2019). In studies with human subjects, researchers have used fMRI scans and high-density electrocorticography signals to accurately decode mental imagery and silent speech (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013; Kay, Naselaris, Prenger, & Gallant, 2008). Recent work on intracranial EEG recordings of speech-related brain activity has achieved remarkable accuracy in identifying brain activity patterns related to inner speech (Moses, Leonard, Makin, & Chang, 2019) while ML techniques have helped enhance the analysis of cognitive processes also from EEG measurements (Hubbard, Kikumoto, & Mayr, 2019; Omurtag, Aghajani, & Keles, 2017). Further, fMRI can show areas of increased activity associated with memory processes. A relatively recent analysis technique based on intelligent algorithms, called multi-voxel pattern analysis (MVPA), has shown potential to detect distinctive patterns of brain activity that appear when people remember particular experiences, hence to decode memory-related information. MVPA is a statistical method in which fMRI data are processed by a computer algorithm that automatically learns the neural patterns associated with specific thoughts or experiences. One of the first applications of MVPA to human memory was published in the journal Science in 2005 (Polyn, Natu, Cohen, & Norman, 2005). In that study, researchers at the University of Pennsylvania and Princeton had subjects view a series of images of famous people, places, and common objects while undergoing fMRI. Using the data collected during this period, the researchers trained an intelligent algorithm to identify activity patterns associated with each of the three categories (famous people, places or objects). Subsequently, subjects were asked (again during fMRI) to recall previously seen images. In performing this second tranche of neural scanning, the algorithm proved capable of reconstructing specific neural traces for each category of images as well as predicting them with good accuracy a few seconds before each response. For example, before the subject named a famous person, the algorithm was able to identify the 'famous person-like' activity pattern in the stream of fMRI data, which included activation of an area of the cortex involved in face processing. Similarly, before the subject named a chair, the algorithm was able to detect the activity pattern "similar to an object" in the neural data, and so on.

Over the past fifteen years, algorithmic techniques for decoding memories have been refined and have now acquired some capacity to predict a person's memories from their brain data. In one study, researchers were able to identify, reconstruct and differentiate between the visual memories of people who watched a certain TV series episode. Based on neural data alone, the researchers were able to reveal very detailed mnemonic details such as whether the scene was filmed outdoors or indoors and whether the protagonist was present or not. Although the subjects recalled the same scene using different words, the neural traces in their brains decoded by the algorithm revealed very similar activation patterns (Chen et al., 2017).

In memory decoding, a privileged role is certainly played by visual memory. In 2017, US researchers succeeded in decoding the visual content of images perceived by subjects from brain activity. Using deep learning algorithms, a brain-inspired approach to AI, the researchers built a model of how the human brain encodes information. Subjects were asked to watch hundreds of short videos while an fMRI machine measured neural activity in their visual cortex and some other area of the brain. Meanwhile, an artificial neural network used for image processing was trained to associate video images with brain activity. When the subjects were then asked to watch new videos, the algorithm was able to predict with good accuracy which areas of the brain would be activated and even visualise which features of the visual information were being processed by each area of the cortex. In parallel, another neural network was able to predict with around 50% accuracy what the subject was looking at by selecting one of 15 categories of images. The researchers demonstrated that, thanks to AI, it is possible not only to decipher memories but also to reveal mental images from another person's brain. To do so, they trained a neural network to partially reconstruct the visual content of what a participant has seen, converting brain signals into pixels.

Besides memory, the same techniques described above can also be used to infer other types of mental information. One notable example is **hidden intentions**. Using neurotechnology and AI, various groups of scientists have succeeded in recent years not only in identifying conscious processes, but also what have been termed a person's hidden intentions and, in this way, predicting future actions. Studies of this kind have repeatedly shown that unconscious neural activity chronologically precedes and potentially influences the free decisions of human beings. For example, neuroscientists led by the British psychologist John-Dylan Hynes claimed to be able to decode from the brain activity of some study participants relevant information about the actions they intended to perform. The task in question was to decide whether to add or subtract two numbers and keep their intention secretly hidden for a few seconds. During this short interval, it was possible for the scientists to determine with 70% accuracy which of the two actions (adding or subtracting) the subjects intended to perform secretly (Bles & Haynes, 2008). Similar results have been successfully obtained with respect to motor choices and reasoning choices, respectively, predicted from previous brain signals. For example, researchers at the Bernstein Centre for Computational Neuroscience in Berlin have shown that the outcome of a free decision to add or subtract numbers can already be decoded by neural activity in the medial prefrontal and parietal cortex four seconds before the participant reports that he is consciously making his choice. These predictive choice signals suggest that unconscious preparation of free choices is not limited to motor preparation. Instead, decisions involving very different scales of abstraction appear to be anticipatable from the dynamics of prior brain activity (Soon, He, Bode, & Haynes, 2013). Studies of this kind, all inspired by research conducted in the early 1980s by the physiologist Benjamin Libet, open up scenarios of great uncertainty regarding the concept of **free will** ad all the other ethical-legal concepts that derive from it.

Thanks to ML techniques, brain scans can be used today not only to observe or predict intentions and memories related to binary choices (addition/subtraction) in well-defined experimental contexts. Furthermore, they can also be used to decode more general **preferences**. A US study showed that fMRI scans can be used to successfully infer users' political views by identifying functional differences in the brains of Democrats and Republicans, respectively (Schreiber et al., 2013). Similarly, lifestyle preferences have been correlated with specific functional differences in male versus female brains. Other studies are even training machine learning algorithms in order to predict from neural data very private preferences such as sexual orientation (Safron et al., 2018), desire for illicit substances such as cocaine or pleasure for gambling (Kober et al., 2016).

The possibility of non-invasively identifying such mental correlations from functional brain differences has attracted particular interest outside the biomedical domain, especially for marketing purposes. Already in the early 2000s, McClure and colleagues used fMRI to show functional differences (increased activation in the dorsolateral prefrontal cortex and hippocampus) in the brains of people who consciously drink Coca Cola compared to the same people who drink the same drink without a label. Their results showed that marketing strategies (e.g. the presence of the Coca Cola brand) can lead to different responses in the brains of consumers in a double-blind study (McClure et al., 2004). This means that when we drink a Coca Cola from a bottle in which the brand label is visible, we like that same drink more than when we drink it from a bottle without a label, even though the drink is exactly the same. These results paved the way for the creation of a branch of neuroscience at the intersection with marketing studies called **neuromarketing**. This area of research has expanded rapidly over the last decade. Today, several multinational companies such as Google, Disney, CBS and MacDonald's use neuromarketing research services to measure consumers' preferences and impressions of their advertisements or products. In addition, a number of companies specialising in neuromarketing, including EmSense, Neurosence, MindLab International and Nielsen, routinely apply neuroimaging techniques, mainly fMRI and EEG, but also Steady State Topography (SST) and physiological measures (e.g., galvanic skin response) to study, analyse and predict consumer behaviour.

Even **dreams** have turned out to be, to some extent, algorithmically decodable via neurotechnology. In 2013, researchers at the Computational Neuroscience Laboratory in Kyoto, Japan, demonstrated the ability to decode dreams based solely on the brain activity of sleeping subjects. By developing a neural decoding approach based on machine learning models, the researchers were able to detect, classify, and predict the content of visual images dreamt by sleeping subjects undergoing fMRI (Horikawa et al., 2013).

Memories, mental images, intentions, and dreams are all essential processes of human cognition whose functional dynamics and content are increasingly accessible to empirical study. By combining neurotechnology and artificial intelligence, it is now possible to decode various components of the brain's extremely rich information content. Because of this information potential, these technologies have often been classified under the label of 'brain reading' (Haynes, 2011), based on the analogy between the possibility of decoding information and mental states from neural data and the functional interpretation of a written text through reading. Some of these studies, as we have seen, have managed to achieve a sufficient degree of epistemological soundness to even build predictive models.

In sum, in recent years the quantitative and qualitative richness of neural recordings has been progressively and rapidly improving. This process of improvement has been influenced not only by the hardware enhancement of the machines, but also, and above all, by the improvement of the analysis techniques, an improvement within which the use of artificial intelligence has played and is playing a key role.

1.3. WHAT INFORMATION NEUROTECHNOLOGY CANNOT READ

Current neurotechnologies, especially non-invasive techniques, are not yet able to *decode thoughts*. This means that neurotechnologies, at the current level of technological development, are not capable of providing a full, granular, real-time and propositionally or experientially describable account of the neural patterns of specific mental processes such as memories or emotions. Decoding thoughts would require a

capacity to reveal the content of mental states, called **mental content**. Mental content regards how a certain mental state comes to be about what it is about. For example, the mental content of someone’s memory is what is actually remembered by that person. The mental content of a someone’s perception is what is actually perceived by that person etc. In the philosophy of mind, it is usually assumed that mental contents can be of two main types: conceptual and non-conceptual.
Conceptual content is the **semantic content** of a mental state, which is believed to be analogous to the content that we can find in words, expressions and sentences of a verbal language. Mental states with conceptual content are also called «propositional attitudes». The conceptual content of a mental state, say a memory, is a particular proposition that may be in principle expressed by a sentence. For example, if a person is experiencing an episodic memory of a past event, such as that “the night of her 18^{th} birthday her mother was wearing a white blouse”, decoding the semantic content of a mental state would require to build a neurotechnology that is capable to identify in the brain signals of that reminiscing person the neural correlates of a proposition stating that the night of her 18^{th} birthday her mother was wearing a white blouse. Ideally, such neurotechnology should also be capable of reconstructing the proposition “the night of my18^{th} birthday my mother was wearing a white blouse” from the brain data.

Non-conceptual content, in contrast, is mental content that is not expressed in the form of a proposition but is experiential, qualitative or phenomenological in character. According to a long tradition of philosophical and scientific thought, non-conceptual content is the content that feelings, experiences, and sensations are typically believed to have. Whereas conceptual content is semantically evaluable in a quite direct way, non-conceptual content is not so. Mental states with non-conceptual content are usually called “qualitative”, “experiential” or “phenomenal” states. The content of these mental states is of qualitative, experiential, or phenomenal character, hence not identifiable with any proposition. For example, if a person is having a certain qualitative experience while smelling the scent of a wisteria flower, decoding this non-conceptual content would require recognizing the neural correlates of this specific qualitative experience (and not of other experiences) in the person's neural data. Ideally, such qualitative experience identified via neurotechnology should be, at least in principle, replicable also in the brain of another person. Once the neural correlates of that exact experience have been identified, recreating those same neural correlates in the brain of another person should in principle induce the very same qualitative experience in another person.

At the current stage of neurotechnology development in 2021, neither conceptual nor non-conceptual mental content decoding is possible. However, some recent studies are showing that bits of conceptual content can be decoded from human brain activity. Jack Gallant and his team at the University of Berkley analysed how the brains of seven subjects responded, under fMRI, to the vision of 129 min of natural movies drawn from movie trailers and other sources (Huth et al., 2016). They created a decoding algorithm that proved capable of decoding detailed information about the object and action categories present in natural movies from human brain activity signals measured via fMRI. Although this algorithm is not sufficiently powerful to decode the full semantic content of mental information, it is sufficiently powerful to accurately decode the presence or absence of general semantic categories (e.g., animal vs structure), specific categories (e.g., canine, wall), and actions (e.g., talking vs running). In other words, the algorithm is not capable of revealing a mental representation such as “this is my beloved pet Fuffy” but it may be capable of decoding the mental content “this is a dog” or even “this dog is running”.

A realistic assessment of the current limitations of neurotechnology-enabled mental decoding is necessary to avoid unrealistic public expectations and guide evidence-based governance. It should be highlighted, however, that current limitations of neurotechnology in mental decoding are contingent, not necessary. Since mental states result from and are built by neuronal activity, there is no logical reason why mental content will forever remain undecodable. As both the hardware and software of neurotechnology improve, the ability of neurotechnology systems to decode mental content will improve accordingly.

Besides current limitations in the decoding of mental content, it should also be highlighted that many neurotechnologies currently available in the consumer space have limited precision. Many manufacturers of consumer neurotechnology products have been observed to advertise market claims (e.g. improving mental wellbeing) that are either unsubstantiated or only loosely corroborated by scientific evidence (Wexler & Reiner, 2019). Again, however, the limited precision of current consumer products is a contingent limitation, possibly caused by the relatively new stage of development of this industry sector and the relatively little validation of those systems. As consumer neurotechnology companies improve their hardware and software, and more brain data become available for analysis and algorithm training, it is plausible to predict that both the accuracy and epistemic power of consumer neurotechnology products will increase accordingly.

2. ETHICAL CHALLENGES

2.1. NEUROETHICS AND NEUROLAW

Since the 1990s, the continuous development of neurotechnology and its growing application in the biomedical sector has elicited ample reflection on the ethical and legal implications associated with the exploration and/or alteration of the human brain. As a result of these reflections, two new fields of normative inquiry have emerged at the intersection of neuroscience, bioethics, medicine, and law: neuroethics and neurolaw.

The word “neuroethics” was coined by William Safire in 2002 and originally defined as "the examination of what is right and wrong, good and bad about the treatment of, perfection of, or unwelcome invasion of and worrisome manipulation of the human brain"(Safire, 2002). In the same year, Adina Roskies proposed dividing the field into two intimately related branches: the ethics of neuroscience and the neuroscience of ethics. She defined the former as a moral framework aimed at regulating, ordering, and guiding behavior in neuroscientific research. In contrast, she defined the latter as the empirical study of how morality as such originates in and is realized by the human brain (Roskies, 2002). In this report, we will focus almost exclusively on the ethics of neuroscience, albeit some reflections are equally valuable from the perspective of the neuroscience of ethics.

The origin of the term “neurolaw” is about a decade older than ‘neuroethics’ as it was first coined by J. Sherrod Taylor et al. in 1991 to denote the growing area of collaboration between neuropsychologists and lawyers in the criminal justice system (Taylor, Harp, & Elliott, 1991). However, this was a very narrow denotation which could hardly encapsulate the bandwidth of modern neurolegal studies. In later years, this meaning was expanded to denote the whole area of intersection between neuroscience and the law (Shen, 2016).

For both neuroethics and neurolaw, an historical milestone was marked, in May 2006, by the foundation of the International Neuroethics Society (INS), which was the by-product of a meeting held in Asilomar, California, in the same year. The INS played a pivotal role towards the institutionalization of neuroethics and neurolaw as academic disciplines.

Throughout the 1990s and early 2000s, four main thematic families dominated public opinion and academic reflection on neuroethics and neurolaw: the ethical permissibility of cognitive enhancement (especially via nootropics)(Farah et al., 2004; Turner & Sahakian, 2006); the philosophical-legal implications of the neuroscience of free will with special focus on the notions of moral responsibility and legal culpability (Fins, 2004; Moreno, 2003; Pereboom & Caruso, 2002); the ethics of neuroimaging and mind reading (Farah, 2002) (Illes, Kirschen, & Gabrieli, 2003; Illes et al., 2004); and the validity and permissibility of neuroscientific evidence in court (Moreno, 2003; Reider, 1998; Zeki, Goodenough, & O'Hara, 2004).

Since the early 2000s, a fifth and complementary area of neuroethical and neurolegal investigation emerged, which started looking at ethical-legal challenges in neuroscience and neurotechnology in terms of high-level normative principles such as rights, duties, and entitlements (see 3.2). This way of analyzing the ethical and legal implications of neuroscience has come to be known as “neurorights”.

The ethical implications of neuroscience and neurotechnology can be grouped into three main thematic clusters: privacy; autonomy, agency and responsibility; justice. In the following, I will provide a concise analysis of each of these clusters.

2.2. PRIVACY

Privacy is a primary ethical concern related to the collection, sharing and processing of brain data. While challenges to privacy arise from the processing of any human data, it is believed that the processing of brain data raises new challenges to the notion of privacy for four main reasons: limited conscious control over one's own brain recordings, protection of the locus internus, informational richness, and risk of neurodiscrimination.

First of all, privacy is both a right and an ability. As such, it relies on an individual's conscious ability to filter the flow of data and intentionally isolate private information. Brain data, in contrast, are mostly elusive to conscious control, hence cannot always be intentionally secluded. The types of information potentially accessible through neurotechnology include not only conscious brain processing but also subconscious processing (e.g., unconscious cognition or subconscious affective states), over which an individual has, by definition, limited or no conscious control at all. For example, when a participant in a neuroimaging study provides her consent to have her brain activity recorded for the study objective stated in the informed consent form, it is still possible in principle to collect redundant subconscious information without their awareness or authorisation. It could be argued that when one person consents to allow the acquisition of brain data, they agree to surrender the protection of their mental information, at least to some extent.

However, in scenarios where brain data collection is either mandated (e.g., in the military sector or workplace) or competitively advantageous (e.g., Facebook’s plan to make brain-typing faster than the touchscreen), the risk of sharing data under explicit or implicit coercion is concrete. Furthermore, it is possible to record redundant data, that is data related to a brain function and/or structure and for a purpose other than the one to which an individual has explicitly consented. This increases the risk of violating the requirement of *purpose limitation*, namely the principle that data collected for one specified purpose should not be used for a new, incompatible purpose. This requirement is protected under the EU General Data Protection Regulation (GDPR)[1] which states that personal data be collected for specified, explicit, and legitimate purposes, and not be processed further in a manner incompatible with those purposes (Article 5(1)(b), GDPR).

While the problem of limited conscious control is shared with other data types (e.g., genetic data), it acquires greater ethical complexity in the neural domain. Specifically, brain data admit no separation between the processed data and the system that generates those data and makes decisions about their processing (the human brain). Ienca and Andorno have called this the “inception problem” (Ienca & Andorno, 2017b). Further, brain data, unlike other physiological measurements, can be argued to have semantic content, hence being propositional in nature. As we have seen earlier, a moderate degree of semantic decoding (in particular, decoding of conceptual categories) has already been proven possible by neuroimaging studies (Huth et al., 2016). Therefore, the unauthorized acquisition of a person’s brain data, for example as a form of mandated evidence in court or by law enforcement during interrogation, may violate the right to silence and the privilege against self-incrimination, i.e., a legal principle that guarantees any individual the right to refuse to answer questions from law enforcement officers or court officials. In criminal investigations, the property of brain activity to encode semantic content and propositional attitudes raises the uncertainty of whether it should be considered physical or testimonial evidence. The first type of evidence consists of items such as hair, blood samples, fingerprints, and other biological materials. The second consists of statements or words spoken by the

[1] The reflections on privacy and data protection contained in this report are primarily focused on the EU General Data Protection Regulation (GDPR), since it is one of the most comprehensive data protection laws in the world, having also an extraterritorial impact on other legal systems.

defendant, a victim, or witnesses. Evidence that brain data can be reliably used as testimonial evidence in court is currently scarce.

Second, brain information pertains to what the XVII Century philosopher René Descartes called the '*locus internus*', that is a person's internal place (also called 'forum internum'). This internal place includes unspoken information, preconscious preferences, memories, attitudes, hidden intentions, and beliefs. These types of information are internal because they exist in a person's mind, possibly in a propositional form, even if that person does not externalize them into the outside world via speech, writing or other behaviour. As such, it has been argued that brain information is the last resort of informational privacy since it includes unexecuted behaviour, inner speech or other non-externalized action (Ienca & Andorno, 2017b). In principle, the privacy of the mind can be preserved even if individual behaviour is constantly surveilled through activity tracking, personal digital technology, self-quantification or simple observation. Collecting and processing human brain data, in contrast, allows a certain degree of access to mental information even in the absence of observable behaviour, hence may challenge the privacy of the mind or **mental privacy**.

Third, the **informational richness** of brain recordings means that they may contain in encoded form highly private information about the individuals from whom they are obtained, including some predictive features of their health, attitudes, and mental states. As we have seen before, numerous studies conducted since the beginning of the third millennium have inferred the possibility of decoding mental contents such as hidden intentions (Haynes et al., 2007), concealed information (Bles & Haynes, 2008), natural images (Kay et al., 2008), visual experiences (Nishimoto et al., 2011) and the unconscious generation of free decisions (Bode et al., 2011) from a person's neural data (recorded via EEG, fMRI or other technologies). Some studies have managed to achieve a sufficient degree of epistemological robustness to build predictive models. For example, two famous studies have used fMRI to pre-tell the flow of consciousness (Haynes & Rees, 2005) and a person's choices regarding not only motor preparation but also, surprisingly, abstract intentions. Such studies have raised legitimate clamour in the scientific community as consciousness, intentionality and free choice are essential components of the faculty that is usually referred to as 'free will' in the theological-philosophical tradition. Although the debate is still open as to whether it is possible to decode not only the neural correlates of mental information but also their actual contents, it is undeniable that in recent years the quantitative and qualitative richness of neural recordings has been progressively and rapidly improving. This process of improvement has been influenced not only by the hardware enhancement of the machines, but also and above all by the improvement of the analysis techniques, an improvement within the scope of the research in which the use of artificial intelligence has played and is playing a key role. The decoding of private mental information is expected to become increasingly possible in the near future due to coordinated advances in sensor technology, spatial resolution of recordings, and machine learning techniques for pattern recognition and feature extraction.

In virtue of their informational richness, brain data also have **biometric** aspects. Brain signals make it possible to distinguish or trace the identity of an individual and are potentially linkable to that individual. Some brain recordings (e.g., recorded EEG signals) can be used as a unique biometric identifier, similar to fingerprints or DNA. In 2007, British computer scientists developed an EEG-based biometric framework for automatic identity verification (Palaniappan & Mandic, 2007). Since then, many non-intrusive EEG-based biometric systems have been developed for individual recognition, authentication, and identification of people. However, unlike other identifiable information, brain waves can potentially be recorded for biometric purposes without the individual's awareness (e.g., as a secondary activity in a neuromarketing study), and thus in the absence of a real ability of the individual to consent to the collection and use of such information. As the market for portable EEG-based devices grows, and in the absence of any real ability to obtain informed consent for the processing of the data they generate, new protective responses to the processing of brain data must be established.

Finally, the processing of brain data, especially neuroimaging biomarkers, generates a risk of "**neurodiscrimination**", i.e., discrimination based on a person's neural signatures (indicating, for example, a dementia predisposition), or mental health, personality traits, cognitive performance, intentions and emotional states. Neurodiscrimination is by many respects similar to genetic discrimination, which occurs when people are treated differently by their employer or insurance company because they have a gene mutation that causes or increases the risk of an inherited disorder. Just like the risk of genetic discrimination is an increasingly pressing concern due to the growing availability of genetic testing, including via direct-to-consumer genetic testing (Chapman, Mehta, Parent, & Caplan, 2019), neurodiscrimination will likely become an increasingly pressing ethical concern due to the growing availability of pervasive neurotechnology, including via direct-to-consumer neurodevices. Genetic discrimination is explicitly prohibited under Article 11 (Chapter IV) of the Oviedo Convention, which states that "any form of discrimination against a person on grounds of his or her genetic heritage is prohibited". It may be necessary to include a similar provision to prohibit discrimination against a person because of his or her neurobiological characteristics. Such a provision can help enhance the protection of the dignity of all human beings and promote their rights and freedoms without discrimination (Article 1). If our society values access to health care for the healthy as well as the sick, the neurodominant and the neurodiverse, we should support strict and broad prohibitions against neurodiscrimination in the context of health insurance, including employer-based health insurance, because they may undermine the ethical principles of universal health care and equity in health care systems.

Privacy challenges are also raised by the data management practices of the neurotechnology system or network. From a data management perspective, data subjects may lose control over their brain data in several ways:

(i) by consenting to the collection of their data without being conscious and adequately informed

(ii) by providing informed consent to the processing of their data for a certain purpose but remaining unaware of further reuses of their data for different purposes (including scraping by third parties);

(iii) by being coerced to have their brain data collected (e.g., via employer's mandate or in an interrogation context);

(iv) via unauthorized access to data by third parties;

(v) as a consequence of data theft.

Ienca and Haselager (2016) reviewed the various security challenges of BCIs and identified several vulnerabilities that could be used by malevolent actors to gain unauthorized access to brain-related information through the BCI channel.

2.3. AUTONOMY, AGENCY & RESPONSIBILITY

The increasing use of machine learning and, more generally, of artificial intelligence to optimise the functioning of BCIs also has implications for the ethical notions of autonomy, agency and responsibility. For example, Haselager (2013) hypothesised that when BCI control is partly dependent on intelligent algorithmic components, it may become difficult to discern whether the resulting behavioural output was actually performed by the user. This difficulty introduces a principle of indeterminacy within the cognitive process that starts from the conception of an action (or intention) to its execution, with consequent uncertainty in the attribution of responsibility to the author of this action.

This principle of indeterminacy could call into question the notion of individual responsibility, with obvious repercussions in terms of criminal law and insurance. In addition, it could generate a sense of alienation in the user, the ethical relevance of which is all the greater in the case of a vulnerable individual such as a neurological patient. For example, imagine a patient suffering from tetraplegia using a BCI which is strongly enhanced by intelligent components for the extraction, decoding and classification of information: how will it be possible to determine which components of the patient's actions are attributable to the patient's volition and which to the AI? This question becomes particularly controversial, as mentioned above, in circumstances where the attribution of responsibility has legal significance. More generally, Thompson has observed that the use of BCIs is problematic for criminal law. The reason for that stems from the fact that criminal law requires that someone can only be found criminally responsible if they have satisfied the actus reus requirement: «that the agent has performed some (suitably specified) conduct». In contrast, agents who affect the world using brain-computer interfaces do not obviously perform any conduct, so when they commit crimes using BCIs it is unclear how they have satisfied actus reus (Thompson, 2019).

In addition, there is a possibility that the centrality of such intelligent components in the functioning of BCIs may affect subjective experience, and thus personal identity. This hypothesis has recently gained preliminary empirical confirmation in a qualitative study about the personal experience of patient-users of BCI (Gilbert, Cook, O'Brien, & Illes, 2019; Gilbert, Goddard, Viaña, Carter, & Horne, 2017). It should be noted, however, that while AI may blur subjective aspects of personal identity, an AI-optimised BCI, taken as a whole, can greatly enhance the user's performative ability to act in a given environment, especially when used for motor control by a patient with severe motor impairment. Therefore, it is difficult to determine in an absolute sense whether intelligent BCIs can increase the autonomy of the user. On the contrary, it is necessary to assess case by case and determine under which circumstances, in which time intervals, and in relation to which mental or physical domains a change (positive or negative) in the autonomy of the user is detectable. In carrying out such assessments, it is important to acquire not only quantitative and objective information (e.g., on mathematical measurements or behavioural observations) but also qualitative and subjective information. The latter category includes the user's introspective self-assessments, which are considered a window of access to the first-person phenomenological dimension of the user (Ferretti & Ienca, 2018).

As we have seen, with the increase in non-clinical uses of BCIs, a further ethical challenge will soon be cognitive enhancement or other neuroenhancement, i.e., any functional augmentation of the nervous system. While clinical applications of BCIs are aimed at restoring motor or cognitive function in people with physical or cognitive impairments such as stroke survivors, neuroenhancement applications may, in the near future, produce superior cognitive or physical performance compared to baseline among healthy individuals. This will make it urgent to discuss which types of enhancement are permissible and under which circumstances. Already today, there is a large ecosystem of private companies that market non-invasive BCI to an ever-increasing number of healthy users for purposes such as self- quantification, cognitive training, neurogaming (the use of brain-controlled video games for recreational or competitive purposes), and polysomnography. Some companies, including Emotiv, based in San Francisco, California, publicly claim (albeit without solid scientific evidence) to be able to "help boost wellness and productivity" of cognitively healthy users. Moreover, BCIs for motor control already allow not only the amplification of existing capabilities, but even the acquisition of faculties otherwise not present in human beings: first of all, the telepathic control of robotic devices such as drones and other semi-autonomous vehicles. This mode of human-machine interaction is widely pursued in the transport industry and military sector. Soon, the diffusion of such applications and the scientific corroboration of their functioning mechanisms will make the ethical issue of enhancement unavoidable.

Finally, implications for autonomy, agency and responsibility are raised by the malevolent misuse of neurotechnology by third parties, especially by external interventions that hijack control over a person's neurotechnological systems. It has been experimentally demonstrated that such neurotechnologies can be hacked by malicious actors in order to hijack their control; a mode of attack that could have deleterious consequences for the victim, including the unauthorised extraction of mental information, the expropriation of the victim from conscious control over their robotic limbs or even serious physical and psychological injuries resulting from the intentional increase in the intensity of neurostimulation by third parties to the detriment of the patient (Chaudhary & Agrawal, 2018; Ienca & Haselager, 2016; Pugh, Pycroft, Sandberg, Aziz, & Savulescu, 2018; Pycroft et al., 2016).

2.4. JUSTICE

The justice implications of neurotechnology relate to determining the conditions for just and equitable access to the benefits of neurotechnology. Ienca (2018) has called 'democratization' of neurotechnology any governance approach designed to universalize and evenly distribute the potential benefits of neurotechnology (Ienca, 2019a). The even and equitable distribution of the potential benefits of neurotechnology is of primary relevance for medical neurotechnologies such as DBS implants for Parkinson's disease, intelligent assistive technologies for dementia, neural interfaces that compensate for loss motor function and neurorehabilitation technologies. In recent years, however, authors have started reflecting on the justice and equity implications of neurotechnologies for cognitive enhancement (Ienca, Shaw, & Elger, 2019).

The problem of global disparities in access to neurotechnology for medical use is another key ethical issue from a global justice perspective as many low-income nations in the global south are lacking sufficient diagnostic and therapeutic infrastructures (Palk et al. 2020).

3. HUMAN RIGHTS CHALLENGES

As neurotechnology advances and opens up new opportunities for monitoring and controlling cognitive, affective, conative and physical functions, there is uncertainty as to whether and how the law should cope with such advances. In particular, it remains debatable whether emerging trends in neurotechnology can be entirely addressed at the level of ethical guidelines and self-governance by neurotechnology actors. In recent years, several experts have argued that the complexity of the ethical challenges raised by neurotechnology (and AI) cannot be addressed exclusively via professional guidelines, best practices, and self-regulation. In contrast, it has been argued, it will require a revision or even a radical reform of existing legal concepts at various levels, including civil law, commercial law, criminal law, and philosophy of law. While the scholarly literature devoted significant attention to the emerging applications of neurotechnology in the context of criminal law or the increasing use of neuroscience evidence in courts, little attention has been paid to the implications of neuroscience and neurotechnology for human rights law. This neglected component of the neuroethical-neurolaw discourse is of particular relevance as the universal nature of the human rights framework could provide a solid basis for what Boire called a “jurisprudence of the mind”(Boire, 2001).

Since brain function and mental faculties intersect several domains of human activity, it is unlikely that a one-size-fits-all approach to neurotechnology governance can be effective. Therefore, a framework for global governance of neurotechnology should operate at multiple levels: neuroethics, soft law, responsible innovation, and binding regulation. Most importantly, given the centrality of the brain in human life, the normative challenges of neurotechnology and AI should be grounded on human rights frameworks.

It is beyond the scope of this report to discuss the different theories of the foundations of human rights, or to take a position on them. For the purposes of this report, a broad practical conception of human rights, such as that proposed by Beitz (2011), is adopted. Beitz argues that human rights are "requirements whose object is to protect urgent individual interests against the foreseeable dangers ('standard threats') to which they are vulnerable under the typical circumstances of life in a modern world order composed of states"(Beitz, 2011, p. 109). In general terms, it can be said that the purpose of human rights is to ensure both the necessary negative and positive prerequisites for leading a minimally good life (Fagan 2015).

The ethical challenges posed by BCI and other neurotechnologies prompt us to address a fundamental ethical-socio-legal question: *determining whether, or under what conditions, it is legitimate to access or interfere with a person's neural activity* (Ienca, 2017). This question needs to be asked at various levels of jurisdiction, including at the level of fundamental human rights (Ienca & Andorno, 2017). The reason for this stems from a triple fact: first, as we have seen, neural activity is scientifically explained as the critical substrate of personal identity and, therefore, of moral and legal responsibility. Therefore, the reading and manipulation of neural activity by means of neurotechnological techniques could logically have unprecedented repercussions on the personal identity of users and introduce an element of obfuscation or even indeterminacy in the attribution of moral and legal responsibility. Secondly, brain activity is detectable from every human being regardless of gender, sex, nation, ethnicity, political or religious affiliation. Finally, as noted above, brain data encode not only electrophysiological information but also mental information. According to a tradition going from Homer to John Milton and up to Virginia Woolf, this domain of the mind and the information associated with it, should be protected as the private domain *par excellence*, the last territory inaccessible to the erosion of private information and the rampant intrusion of the data society.

While neurotechnology has the potential to affect human rights such as privacy, freedom of thought, the right to mental integrity, freedom from discrimination, the right to a fair trial, or the principle against self-incrimination, most provisions of international human rights law do not make any explicit reference to neuroscience and/or neurotechnology. In contrast to other biomedical developments, which have already been the subject of national and international standard-setting efforts, neurotechnology still remains largely terra incognita for human rights law. However, the implications raised by neuroscience and neurotechnology for the inherent characteristics of human beings call for a prompt and adaptable response from human rights law.

3.1. LACUNAE IN INTERNATIONAL HUMAN RIGHTS FRAMEWORKS

Human rights instruments, such as the United Nations Universal Declaration of Human Rights (UDHR, 1948) and the International Covenant on Civil and Political Rights (ICCPR, 1966), were drafted before the human brain became measurable (with the exception of very few techniques such as the EEG), amenable to computational analysis and directly alterable via neurofeedback or neurostimulation. Thus, these instruments did not explicitly define the requirements for gaining access to the human brain and the information that can be acquired from it, or for intervening neurotechnologically in the brain in a way that preserves human dignity and human rights (Ienca et al. 2021).

That being said, the UDHR spells out a list of "basic rights and fundamental freedoms" and affirms their universal character as inherent, inalienable, and applicable to all human beings. At a closer look, some of the rights and freedoms contained in the UDHR, do implicitly refer to the protection of the brain and mental sphere (e.g., Article 18 on “freedom of thought”). However, they do not explicitly spell-out technology-specific risk scenarios enabled by neurotechnology or provide a well-articulated framework for protecting the human brain and mind.

In the decades following the adoption of the UDHR, new declarations were drafted with the aim of protecting human rights in light of technological advances. For example, progress in genetics, particularly in genome sequencing and editing technologies, was addressed by UNESCO in the International Soft Law Declaration on Human Genetic Data (2003). This declaration ascribes to human genetic data a “special status” on account of their sensitive nature since “they can be predictive of genetic predispositions concerning individuals and that the power of predictability can be stronger than assessed at the time of deriving the data” (Preamble). Due to this special status, the Declaration defines a set of conditions for legitimate use of human genetic data with special focus on enabling free and informed consent, preventing discrimination and stigmatization, protecting privacy and confidentiality, and ensuring a fair sharing of benefits with society as a whole.

In comparison to genetic data, human brain data remain without explicit guarantees and lack comparable protection by human rights instruments. As another example, the Council of Europe’s Oviedo Convention (1997) contains explicit provisions related to biomedical technologies and practices such as predictive genetic testing, genetic engineering, medically assisted procreation, research on embryos in vitro, as well as organ and tissue removal from living donors for transplantation purposes. However, it contains no provision related to neurotechnology and makes no reference to the protection of the human brain and mind. Moreover, in the international humanitarian law landscape there is no specific international treaty addressing the dual-use or potential weaponization of neurotechnology for military purposes. This is despite the fact that disarmament treaties such as the 1972 Biological Weapons Convention

(BWC), deal with the weaponization of other biological knowledge and technologies such as biological and toxin weapons by prohibiting their development, production, acquisition, transfer, stockpiling and use.

Therefore, addressing the issue of human rights in relation to neurotechnology —and, more broadly, in relation to the human brain and mind— appears to be an increasingly fundamental challenge. In 2017, Ienca and Andorno conducted an ethical-legal assesment of human rights in the age of neuroscience and neurotechnology (Ienca & Andorno, 2017b). Those authors conducted a parallel and comparative analysis of, respectively, emerging trends in neurotechnology and human rights provisions related to the protection of the human brain and mind contained in existing human rights instruments such as the United Nation's UDHR (1948), the European Union's Charter of Fundamental Rights (2000) and the UNESCO's Universal Declaration on Bioethics and Human Rights (2005). Their analysis concluded that existing human rights are necessary but may not be normatively sufficient—or, at least, not sufficiently agile—to respond to the emerging issues raised by neurotechnology. For this reason, the authors argued that «the possibilities opened up by neurotechnological developments and their application to various aspects of human life will force a reconceptualization of certain human rights, or even the creation of new rights to protect people from potential harm» (Ienca and Andorno 2017b).

Normative lacunae have been recognized in both supranational and international law. Ienca et al. (2021) have provided an overview of current gaps in the current legal framework. The most obvious consideration is that no mandatory governance framework focused on the human brain and the information derived therefrom currently exists in supranational or international law. All data generated from neurotechnology systems should be prima facie considered personal data, as defined by instruments such as the legally binding European Union's GDPR, the non-binding 2013 OECD's Privacy Guidelines and the Council of Europe's *Modernized Convention for the Protection of Individuals with Regard to the Processing of Personal Data*. Under these instruments, personal data are defined as any information related to an identified or identifiable natural person (Art. 4 GDPR; Art. 1 OECD Privacy Guidelines, Art. 2a CoE).

However, there are several caveats with this definition in the context of neurotechnology. First, data protection instruments such as the GDPR may not be applicable if the data obtained from neurotechnology systems are anonymized (Ienca et al. 2021). This is despite the fact that the technical difficulty of anonymizing brain data leaves open the potential for re-identification. The possibility of previously de-identified brain data becoming identifiable again is real because of the technologies involved in processing brain data and their high informational richness and contextualization. Researchers have demonstrated that it is experimentally feasible to re-identify data subjects based on electrophysiological measurements or neuroimaging data, and even predict present emotional states and future behaviour from brain data, as well as decode sensitive information from either the neural activity of data subjects or their digital phenotypes (Omurtag et al., 2017; Schwarz et al., 2019).

Article 4, no. 15 of the GDPR also defines a specific type of personal data called health data, which is defined as “data related to the physical or mental health of a natural person.” Prima facie, the information derived from someone’s brain and mind constitutes health data because it can be used to infer the physical and mental health of a natural person. This interpretation is supported by the Article 29 Working Party which further specified that the category of health data includes information about a person’s intellectual or emotional capacity (Kohnstamm 2011). This would imply that brain data are to be afforded some of the highest protections (cf. article 9, GDPR). However, it is unclear whether all brain data can be considered health data and be treated as sensitive, as required by the GDPR. Specifically, it has been noted that brain data generated by consumer neurotechnologies may not currently constitute "health data," and thus may be subject to lesser protections than data from clinical applications, because the application of these devices does not fall under medical device regulatory regimes (Rainey et al., 2020). This means that brain data collected through non-clinical neurotechnologies such as consumer BCIs are underprotected and possible violations of mental privacy and risks of neurodiscrimination may occur.

Second, it is exceptionally difficult to safeguard the data subject's right to be forgotten when it comes to neurotechnology users. The right to be forgotten is the right of someone to request a data controller to remove their personal data from their data directories. As we have seen before, brain data often escape conscious control and can be easily re-identified even when previously deidentified. Therefore, data subjects may not be aware of what data is being collected from them and for what purposes. Moreover, even if a person is initially able to consent to the processing of data and later able to have it deleted, the data controller or others may still use that deleted data to derivatively reconnect the data to the data subject. In the case of brain data involving unconscious information, the data controller might be able to retain data of which the individual is unaware (Ienca et al. 2021). The nature of brain data could also undermine the ability of data subjects to exercise their rights to access, modify, and delete their data. For example, Greenberg noted that a data subject might not own a computer powerful enough to process data from a BCI (Greenberg, 2019). Similarly, deletion of brain data could substantially decrease the accuracy of ML models generated with this data. This implies that the right to be forgotten might be difficult to apply to BCIs because of the impact such data deletion would have on the accuracy of AI-driven predictive models (ibid.).

Third, although international data protection laws such as the GDPR are vital to the protection of personal data, they may offer insufficient protection for brain data because these laws allow for some exemptions to data subjects' rights when data is processed for research or statistical purposes (as defined in Article 5 (1) (b) GDPR).

This is also the case when the research is conducted by a private entity such as a consumer neurotechnology company. Since brain data are largely processed by processors that appeal to scientific and statistical purposes, this implies that some processing of brain data by both public and private actors (e.g., government agencies or consumer neurotechnology companies), may rely on exemptions to key data protection rules. Indeed, it is doubtful under what conditions the privilege to use brain data for scientific research applies to brain data collected outside the biomedical field. For this reason, Ienca et al. (2021) called for greater transparency with respect to the various purposes of research and to allow data subjects to intervene in further processing if it is for undesirable purposes (Ienca et al. 2021).

Fourth, as we have previously seen, brain data may undermine another normative principle strongly associated with information privacy and personal autonomy, namely *purpose limitation*. Sensitive categories of personal data (including health data) can only be collected for specific purposes that must be specified at the time the data subject provides their consent. Under the GDPR, for example, purpose limitation is a requirement that personal data be collected for specific, explicit, and legitimate purposes, and not further processed in a manner incompatible with those purposes (Article 5(1)(b), GDPR). However, implementing the purpose limitation requirement for brain data is made particularly difficult by the fact that current neurotechnologies cannot pre-emptively discern purpose-specific data from the myriad of brain signals recorded by the device (signal-to-noise problem), including subconscious processes. This implies that a large amount of redundant information can be collected of which the data subject is to a large extent unaware. Data security measures such as differential privacy have been proposed to balance consent to broad processing purposes. However, these measures are difficult to define at the statutory threshold of re-identifiability (Ienca et al. 2021). In addition, technical tools for selective filtering such as the Brain-Computer Interface Anonymizer (Chizeck & Bonaci, 2014) and Radio Frequency Identification RFID-based systems for identifying brain activities in a secure real-time mode (Ajrawi et al. 2021) are in the early stages of development. In view of these factors, it may be more difficult for people to exercise control over their brain data compared to other categories of data.

Another area of faulty legal coverage related to neurotechnology is the prevention of neurodiscrimination. Although "all forms of discrimination based on a person's genetic heritage" are prohibited by the Council of Europe Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine (Oviedo Convention), no explicit protection prevents neurodiscrimination. This is problematic because the potential for brain data to reveal sensitive characteristics through processing is very high (Rainey, Bublitz, Maslen, & Thornton, 2019), making it possible to discriminate people based on neurological and/or psychological traits.

Finally, Ienca et al. (2021) have argued that the safeguards provided by data protection regulations may not be adequately scalable to group-level data. This failure to adequately scale protection raises a twofold risk to group privacy: (A), third parties may make inferences about a group of data subjects based on one or more characteristics inherent in their brain data and shared by all individuals in the group. For example, this could be a slower-than-average reaction time to cognitive tests or an increase in brain activity associated with certain patterns of online behaviour. Second, individuals may be unknowingly identified through their brain data, albeit anonymized, as part of an until then unsuspected group (Ienca et al. 2021). This, again, may cause neurodiscrimination against them.

3.2. INTRODUCING THE NOTION OF "NEURORIGHTS"

As we have seen earlier, neuroethics scholarship focused primarily on topics such as the ethical permissibility of cognitive enhancement via nootropics, the philosophical-legal implications of the neuroscience of free will, the ethics of neuroimaging (especially with regard to mind reading), and the validity and admissibility of neuroscientific evidence in court.

Since the early 2000s, a new area of neuroethical and neurolegal investigation emerged, which started looking at ethical-legal challenges in neuroscience and neurotechnology in terms of high-level normative principles such as rights, duties and entitlements. A pioneering step in this direction was marked by Boire's and Sententia's work on the notion of 'cognitive liberty' (Boire, 2001; Sententia, 2004), which was defined by the latter as "the right and freedom to control one's own consciousness and electrochemical thought process" (Sententia, 2004, p. 227). It should be noted that this field of neuroethical investigation emerged in full continuity with the aforementioned dominant debates in neuroethics and neurolaw. Boire, for example, developed his reflections on cognitive liberty contextually with ongoing debates on the ethics of neuroimaging and mind reading (Boire, 2001). Similarly, Sententia developed her definition and normative analysis of cognitive liberty by taking stock of the ongoing debate on cognitive enhancement within the neuroethics community (Sententia, 2004).

The point of departure of their analyses compared to previous neuroethical scholarship, however, is of normative-theoretical nature. Both authors posited that the concept of cognitive liberty should be interpreted not merely as a neurophilosophical description or a *moral desideratum* but as a "fundamental right" (Sententia, 2004, p. 223). This view is well captured by Sententia's argument that advances in neurotechnology require a high-level analysis that is contextual to "those individual rights embedded in our democratic constitution" and her claim that cognitive liberty "is the necessary substrate for just about every other freedom" (Sententia, 2004, 227). In the 2010s, this right-based view of cognitive liberty was further expanded by Farahany's doctrinal analysis of, respectively, the Fourth Amendment to the United

States Constitution and the Self Incrimination Clause of the Fifth Amendment (Farahany, 2012). Further, it was vigorously reaffirmed by Bublitz's thesis that the use of "mind-interventions outside of the therapeutic contexts" urges the law to recognize cognitive liberty (which he also called 'mental self-determination') as a "basic human right" which "guarantees an individual's sovereignty over her mind" (J.-C. Bublitz, 2013). Although none of the above authors made use of the term, this body of scholarship laid the foundations for the emerging area of enquiry at the intersection of neuroethics and neurolaw that is now increasingly known in the public eye as 'neurorights'. This area of enquiry introduced a new angle from which we can look at the ethical-legal challenges in the mind and brain sciences, namely in terms of rights—including both legal rights and rights in the philosophical sense or moral rights—freedoms, entitlements, and associated obligations.

The term *'neuroright'* was coined by Ienca and Andorno in April 2017 in an ancillary article (Ienca & Andorno, 2017a) to their ethical-legal analysis of human rights in the age of neuroscience and neurotechnology (Ienca & Andorno, 2017b). Their analysis concluded that existing human rights are necessary but may not be normatively sufficient to respond to the emerging issues raised by neurotechnology. For this reason, the authors made the case that «the possibilities opened up by neurotechnological developments and their application to various aspects of human life will force a reconceptualization of certain human rights, or even the creation of new rights to protect people from potential harm» (Ienca and Andorno 2017b). In other words, they presented neurorights as an emerging class of fundamental rights related to the protection of a person's brain and mind.

In particular, they identified four neurorights, that, in their view, may become of great relevance in the coming decades: the right to cognitive liberty (which they interpreted in agreeance with Sententia and Bublitz), the right to mental privacy, the right to mental integrity, and the right to psychological continuity. This article sparked a debate in the public media and academic community, with many authors expressing endorsement of this neurorights proposal but also some dissenting voices. Among others, Cascio endorsed the proposal but questioned whether neurorights should be seen as legal rights of the mind or of the person (Cascio, 2017). Further, he critically discussed the limits of neurorights especially in the case of minors. Around the same time, Pizzetti argued, in a letter to the UNESCO Chair of Bioethics, that the four neurorights identified by Ienca & Andorno may constitute the building blocks of a "Universal Declaration on Neuroscience and Human Rights" (Pizzetti 2017). In contrast, Nawrot criticized the proposal and casted doubts on the potential of neurorights to "reconcile the technological infiltration into our interior castle" (figurative for the human brain and mind) with the concept of "freedom of thought" and the "foundation of a democratic state ruled by law" (Nawrot, 2019). Sommaggio & Mazzocca further investigated the relationship between human rights and cognitive liberty (Sommaggio & Mazzocca, 2020). They concluded that the notion of cognitive liberty provides the necessary conceptual ground for building "a human neuro-rights declaration".

About half a year later, the debate on neurorights was reignited and amplified by an article in *Nature* by a team of twenty-five researchers coordinated by Rafael Yuste and Sara Goering (Yuste et al., 2017). The authors identified four areas of concern associated with neurotechnology and AI, namely privacy and consent, agency and identity, augmentation, and bias. For each of those areas of concern, they argued, "clauses protecting such rights (called neurorights)" should be added to international treaties (ivi). This article was extremely influential in the public opinion. By shifting the focus of the neurorights discourse from ethical-legal analysis to policy advocacy, this proposal exerted a great impact on nation-level legislative reforms, most notably in Chile. While the semantics, theoretical justification and normative demarcation of these rights were not fully addressed in the original short article, this proposal was further elaborated in greater detail a few years later by the same group (Goering et al., 2021) as well as by Yuste, Genser and Hermann (Yuste, Genser, & Herrmann, 2021). Furthermore, Yuste's advocacy work resulted first in the establishment of the Neurorights Initiative at Columbia University—the first institutional think-thank on neurorights— then, in collaboration with European and North American partners, of the Neurorights Network, i.e., the first international network of scholars working on neurorights, whose membership currently extends to four continents and ultimately of the Neurorights Foundation.

3.3. DEFINING "NEURORIGHTS"

In this report, **neurorights** are defined as the ethical, legal, social, or natural principles of freedom or entitlement related to a person's cerebral and mental domain; that is, the fundamental normative rules for the protection and preservation of the human brain and mind. Accordingly, neuroright studies are a subfield of neuroethical and neurolegal inquiry concerned with the ethical, legal, social, or natural principles of freedom or entitlement related to a person's cerebral and mental domain; that is, the fundamental normative rules for the protection and preservation of the human brain and mind. As this definition indicates, neurorights are complex and multifaceted rights which are typically interpreted as both moral rights (i.e., rights in the philosophical sense) and legal rights. The study of neurorights is a privileged avenue of enquiry to reflect on whether neurotechnology-related issues can be sufficiently addressed by the existing human rights framework or whether new human rights pertaining to the neuro-cognitive domain need to be entertained in order to govern neurotechnologies.

3.4. ROOTING "NEURORIGHTS" IN THE HISTORY OF IDEAS

Neurorights did not emerge from thin air. In the history of philosophy and legal-political thought, several conceptual constructs can be identified as historical antecedents and conceptual foundations of neurorights. In particular, three main conceptual families can be recognized: freedom of thought and conscience, the right to privacy and the right to mental integrity. Let us look at them in detail.

3.4.1. Freedom of thought and conscience

The thesis that the human mind and the cognitive processes it enables are free is ubiquitous in the history of ideas. Although independent thinking was culturally discouraged and even legally persecuted throughout most of human history, a cultural tradition aimed at the protection of such fundamental freedom runs through, albeit sometimes under trace, the entire history of thought.

One of the very first records of the idea of freedom of thought dates back to the Maurya Dynasty that ruled a large part of the Indian subcontinent in the third century BC. In particular, in the second half of the century, Indian emperor Ashoka the Great issued edicts promoting respect for “freedom of conscience”(Luzzatti, 2006). During the Athenian Democracy, the process of secularization and the continuous political tensions between democracy and oligarchy, led to the emergence of the term (*parrhēsia*). Although the closest literal translation of this term into modern English would be ‘frankness’, *parrhēsia* was often invoked by democracy supporters as a synonym of “freedom of speech” during times when democracy and oligarchy were perceived as mutually exclusive such as during the Peloponnesian War (431–404 BC).

In the first century BC, Paul of Tarsus reflected on the boundaries of freedom of thought. In his first letter to the Corinthians, he discussed the extent to which someone’s freedom [eleutheria] should be judged by another's conscience [suneideseos] 10:29 (Collins & Harrington, 1999). In Christian philosophy, the notion of freedom of thought was often intertwined with the notion of *liberum arbitrium*, which is typically translated into English as “free will”. However, while freedom of thought and conscience constituted a normative principle (typically related to a political commitment to religious tolerance), free will was originally conceptualized as a descriptive statement about the lack of necessity of human will. This descriptivist definition of free will was rooted in the late ancient Greek philosophy, especially among the Stoics. For example, the stoic philosopher Epictetus (50 – c. 135 AD), considered as a “fact that nothing hindered us from doing or choosing something that made us have control over them" (Long, 2002).

Throughout the Renaissance, several concepts of similar nature emerged. For example, in the 17th Century, Puritan minister and theologian Roger Williams coined the notion of “soul liberty”, namely the idea that God had endowed human beings with the inborn right to make choices in matters of faith (Gaustad, 2001). This notion subsequently evolved into the notion of “freedom of religion” or “religious liberty” which is currently protected under the UDHR, Article 18. This article guarantees the freedom of every human being to change his religion or belief, as well as the freedom to manifest his religion or belief in teaching, practice, worship, and observance, either alone or in community with others and in public or private. The United States Bill of

Rights also contains a guarantee in the First Amendment that laws may not be made that interfere with religion "or prohibiting the free exercise thereof".

The poet John Milton, a coeval of Williams, used the expression “freedom of the mind” to indicate the right and ability of people to protect their minds from external interference (Milton, 1791). Milton was among the first thinkers to introduce the idea that the human mind is the last refuge of personal freedom and self-determination. In the 19^{th} Century, this idea was echoed and further expanded by John Stuart Mill, who argued that “[o]ver himself, over his own body and mind, the individual is sovereign” (Mill, 1859)p. 12). As Martah Farah observed, this individual sovereignty over herself is grounded in “an intuition about individual freedom… not previously denied even to prisoners: the freedom to think one’s own thoughts and have one’s own personality” (Farah, 2002) p. 23). Finally, in the 20th Century, the novelist Virginia Wolf famously reaffirmed this idea as she wrote: “There is no gate, no lock, no bolt that you can set upon the freedom of my mind”(Woolf, 1929). This view of the mind as the ultimate site of personal freedom was highly influential for the debate on neurorights. For example, Sententia echoed this tradition arguing that “the right and freedom to control one’s own consciousness and electrochemical thought processes is the necessary substrate for just about every other freedom” (Sententia, 2004). Similarly, Levy has stated that “if we have the right to a sphere of liberty, within which we are entitled to do as we choose, our minds must be included within that sphere” (Levy, 2007; 38 p. 179)

Freedom of thought in the normative sense is protected under the Universal Declaration of Human Rights (UDHR), which is legally binding on member states of the International Covenant on Civil and Political Rights (ICCPR). In particular, the right to freedom of thought is listed under Article 18, which states the following:

Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest his religion or belief in teaching, practice, worship and observance (Art.18).

Freedom of thought is also protected under Article 9 of the European Convention on Human Rights (ECHR, 1950). However, the ECHR provides a relatively narrow definition of freedom of thought, which intimately ties this freedom to the freedom of conscience and religion. In particular, it explicitly states that “one of the present-day issues of respect for freedom of thought, conscience and religion is embodied, at both international and national level, in the upsurge of religious intolerance» (Art.9).

In contrast, the United Nations Human Rights Committee (UNHRC) has emphasized that the scope of the right to freedom of thought is “far-reaching and profound; it encompasses freedom of thoughts on all matters” (United Nations Human Rights Committee, 1993). In other words, although the UDHR establishes a prima facie link between freedom of thought and freedom of religion, it does not reduce the former to the latter as freedom of thought should be intended as far-reaching and profound in scope. Further, the UNHRC has clarified that the "the freedom of thought, conscience, religion or belief" should be distinguished from "the freedom to manifest religion or belief"(Committee, 1993). Further, it clarifies that the UDHR "does not permit any limitations whatsoever on the freedom of thought and conscience or on the freedom to have or adopt a religion or belief of one's choice. These freedoms are protected unconditionally" (ivi).

In the neurorights debate, Ienca and Andorno have further emphasized this distinction between freedom of thought and the freedom to manifest thought or belief. They argued that cognitive liberty protects the sphere of thought even prior to any externalization or manifestation of thought through speech, writing, or behavior. As such, they argued, cognitive liberty is chronologically antecedent to any other freedom (Ienca & Andorno, 2017b) and complementary to notions such as freedom of speech, freedom of the press and freedom of assembly.

In the United States, the protection of freedom of thought is frequently associated with the First Amendment (Richards, 2015). Although it does not mention freedom of thought explicitly, U.S. courts have explicitly referred to a "First Amendment right to freedom of thought" (Doe v. City of Lafayette, Indiana, 2003) and the U.S. Supreme Court has stated that "at the heart of the First Amendment is the notion that an individual should be free to believe as he will" (Abood v. Detroit Board of Education, 1977).

Freedom of thought has often been considered the precursor and progenitor of other liberties such as freedom of religion and freedom of expression. In virtue of this precursory nature, freedom of thought is axiomatic for many other freedoms, since they are in no way required for it to operate and exist. This fundamental role of freedom of thought as the substrate of other freedoms has been recognized, among others, by the U.S. Supreme Court Justice Benjamin Cardozo whose reasoning in court case Palko v. Connecticut (1937) was the following: «Freedom of thought... is the matrix, the indispensable condition, of nearly every other form of freedom. With rare aberrations a pervasive recognition of this truth can be traced in our history, political and legal» (Polenberg, 1996). Sententia's argument that cognitive liberty should be considered the substratum of all other freedoms can be subsumed into this legal-philosophical tradition. By virtue of this precursor nature, freedom of thought can be considered axiomatic for the other freedoms, since these freedoms are in no way required for it to operate and exist.

While it appears clear that the right to freedom of thought adequately protects externalizations of thought such as observing or changing a religion, it is questionable whether this right is sufficiently broad to also protect thought in the neuropsychological sense. Legal provisions protecting freedom of thought appear well-equipped to protect the "locus externus"—such as behavior, verbal utterances, written text— but less-equipped to protect the “locus internus”—such as unspoken information, silent speech, hidden intentions, preconscious preferences, and attitudes.

Finally, since recent reports indicate that compelled neuromonitoring is occurring among workers at state-backed facilities in China, human rights frameworks should protect the ability of people to make free and competent decisions about the collection and processing of their personal brain data. The European Convention on Human Rights (ECHR), which protects the rights to privacy and freedom of thought (Arts. 8 and 9) may offer the suitable conceptual and normative framework to prevent coercive uses. When the CoE Modernised Convention comes into force, it could serve as a solid basis for further specification. However, further specifications may be needed to expand the scope and focus of the right to freedom of thought beyond the sole (respect for) freedom of conscience and religion, as to protect all expressed and unexpressed cognitive, affective, and other mental states. ECHR’s Article 9 clause, which emphasizes that “all recognized beliefs are protected by this right” may offer a suitable basis for this broader characterization. However, beliefs are generally considered to represent a special class of mental states characterized by semantic content and propositional attitudes. For this reason, mental representations involving sensory qualities such as perceptions and episodic memories are usually not considered to be beliefs. Furthermore, some propositional attitudes such as desires, by their mode or the way how they are directed at propositions: while beliefs try to represent the world as it is and they do not intend to change it, desires generate representations of how the world could or should be. In other words, beliefs are a too narrow category, hence insufficient to protect the whole mind. For this reason, future human rights reforms may be needed to ensure that the right to freedom of thought shall protect not just all beliefs but all mental states.

3.3.2. Privacy

Although the right to privacy was partly encapsulated in the notions of freedom of thought and personal autonomy, the first consistent conceptualization of the modern right to privacy dates to a seminal article, published in 1890, by Samuel Warren and Louis Brandeis. In this article, privacy was conceptualized as “a right to be let alone” (Brandeis & Warren, 1890). At the time their article was written, Warren’s and Brandeis’ primary concern was the increasing interest of the press in gossiping and revealing personal information about individuals without their consent. This specific instance of privacy was further developed by Alan Westin and other authors into the broader notion of «information privacy», i.e., the control over information about oneself. According to Westin, information privacy can be defined as everyone’s claim to determine for themselves when, how, and to what extent personal information is communicated to others (Westin, 1968).

International human rights law formally recognises the right to privacy. The Universal Declaration of Human Rights (UDHR) states that “no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks” (Article 12). Similarly, the 1950 European Convention on Human Rights (ECHR) stipulates that “everyone has the right to respect for his private and family life, his home and correspondence” and specifies that this right involves “protection against telephone tapping, collection of private information by a State’s security services and publications infringing privacy” (Article 8).

In today’s digital world, the right to privacy has become relevant to entire new domains and methods of information processing that were unthinkable at the time of Warren and Brandeis or even of the UDHR and the ECHR; among those: the brain-mind sphere and data processing techniques aimed at revealing information about a person’s mental processes or neurological health. This category includes both the predictive analysis of primary neural data such as brain recordings and inferences based on secondary data (e.g., phenotypic or behavioural data) through techniques such as affective computing. For example, Yuste et al. argued that “an extraordinary level of personal information can already be obtained from people’s data traits” and argued that “citizens should have the ability — and right — to keep their neural data private” (Yuste et al., 2017). Based on similar considerations—with special regard to the security vulnerabilities of neurodevices, the nature of neural data and the inferential potential of advanced data analytic techniques—Ienca and Andorno proposed to evolutionarily reinterpret the right to privacy and proposed the recognition of a “right to mental privacy” which would explicitly protect individuals against the unconsented intrusion by third parties into their mental data (be it brain data or proxy data indicative of neurological, cognitive and/or affective information) as well as against the unauthorized collection of those data (Ienca & Andorno, 2017b).

A curious historical antecedent of the right to mental privacy is reported by the XVI-XVII Century philosopher Francis Bacon, who chronicled that Queen Elisabeth I revoked a thought censorship law in the late sixteenth century, because, allegedly, she did "not [like] to make windows into men's souls and secret thoughts" (Brimacombe, 2000). In the early 20th century, historian John Bagnell Bury has emphasized the relationship between mental privacy and freedom of thought. In his famous “A History of Freedom of Thought”, he argued that “a man can never be hindered from thinking whatever he chooses as long as he conceals what he thinks” (Bury, 1914)(p.1). This suggests that exercising one’s right to mental privacy—and thereby concealing one’s own thoughts—is necessary to fully exercise one’s own right to freedom of thought.

In the light of the historical context in which it was drafted and adopted, it is unsurprising that Article 8 of the ECHR makes explicit reference to telephone tapping but does not refer to privacy challenges in digital technology (digital privacy) and neurotechnology (neuroprivacy and mental privacy). However, as the world of 2021 is profoundly different from the world of 1950 with regard to how information is collected, processed and shared, new provisions may be needed to protect privacy in the context of ever-evolving technology. The Oviedo Convention, in contrast, offers a sufficiently broad and technology-agnostic basis for regulating mental privacy as it states that "everyone has the right to respect for private life in relation to information about his or her health» (Art 10(1)). However, this article of the Convention appears too narrow in scope as it only protects health information, hence may not adequately protect mental or neural data processed for non-health-related purposes such as, for example, information about hidden intentions or personal beliefs.

3.3.3. Mental Integrity

If freedom of thought protects the human brain and mind from undue external interference and privacy rights protect personal information (including mental information) from external intrusion, other normative principles protect the human brain and mind from harm. In the history of ideas, the ethical principle of "non-maleficence" is the most comprehensive conceptual construct that postulates the protection of a person's integrity and the avoidance of harm.

The moral obligation "to abstain from doing harm" (Ancient Greek: ἐπὶ δηλήσει δὲ καὶ ἀδικίῃ εἴρξειν) is already present in some early versions of the Hippocratic Oath and is widely reported throughout the medical deontology literature. In the book "Epidemics" of the Hippocratic Corpus it is stated that: "The physician must ... have two special objects in view with regard to disease, namely, to do good or to do no harm" (book I, sect. 11, trans. Adams, Greek: ἀσκέειν, περὶ τὰ νουσήματα, δύο, ὠφελέειν, ἢ μὴ βλάπτειν). This moral obligation was subsequently rephrased into the Latin maxim "*primum non nocere*", that is "first do no harm"[2]. Although the principle of avoiding harm is entrenched in the ethos of medicine and biomedical research, the characterization of harm is not always semantically unambiguous. The medical ethics literature classifies harm according to its magnitude, severity, duration, and reversibility (Meslin, 1990). Further, it distinguishes various types of harm depending on the personal sphere or capability affected by the malicious intervention. These include physical, psychological, and socio-economic harm. However, the separation of physical and psychological harm is questionable as it implicitly assumes a dualistic ontology of the person (body vs mind). Further, it has been observed that novel forms of harm enabled by emerging technologies may not easily fit into this classification

[2] Unlike usually assumed, the Latin phrase "primum non nocere" is not of ancient origin. Smith (2005) traced it back to an attribution to Thomas Sydenham (1624–1689) in a book by Thomas Inman (1860) entitled *Foundation for a New Theory and Practice of Medicine.* See: *(Smith, 2005)*

(Favaretto, De Clercq, Gaab, & Elger, 2020; Hayes, 2017). The prevention of psychological harm, such as harm from psychological abuse, is one eminent historical antecedent of neurorights, especially of the right to mental integrity.

Historically, the phrases ‘mental integrity’ and ‘psychological integrity’ have been used, albeit unsystematically, to refer to the psychological counterpart of the principle of physical integrity (also called ‘bodily integrity’). Physical integrity refers to the normative principle of inviolability of the human body, as well as the ability and right of the person to exercise autonomous control over their body. As such, physical integrity is a fundamental requirement for self-ownership and self-determination.

According to Martha Nussbaum’s capability approach —a normative approach to human welfare based on the capability of the person to achieve their wellbeing— physical integrity is more than a right because it is also one of the ten principle capabilities. In fact, she defines it as a set of abilities such as the ability “to move freely from place to place” and the ability “to be secure against violent assault” (Nussbaum, M. (2007). Human rights and human capabilities. Harv. Hum. Rts. J., 20, 21).

Physical integrity is currently protected both in many national legislations and in international human law. At the national level, physical integrity has been recognised inter alia by the courts of the Republic of Ireland as an “enumerated right” which is protected by the general guarantee of "personal rights" contained within Article 40 of the Irish constitution. At the international level, both the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights (ICCPR) protect physical integrity.

In contrast, occurrences of the right to mental integrity in the history of ideas are relatively rarer. As Douglas observed, “in contrast to the right to bodily integrity, the putative right to mental integrity enjoys no significant philosophical pedigree” (Douglas, 2014). As a notable exception, Welford used, in the early 1970s, the notion of mental integrity as a demarcating criterion for delimitating the ethical boundary between the obligation to offering life-maintaining treatment and unreasonable therapeutic obstinacy, especially among terminally ill patients, the senile and seriously defective infants (Welford, 1970).

The right to mental integrity—together with physical integrity— is protected under the EU’s Charter of Fundamental Rights (CFR), whose Article 3 states that “everyone has the right to respect for his or her physical and mental integrity.” The Charter focuses in particular on four requirements: free and informed consent, the non-commercialization of body elements, and the prohibition of eugenic practices and human reproductive cloning. Further, it promotes a right of access to mental health services and prevention of eugenic practices. However, no explicit reference is made to neurotechnology-related practices or specific harms caused by malevolently interfering with a person’s neuropsychological sphere. Most importantly, as Douglas noted, “very little work has been done to determine […] what varieties of mental influence fall within its scope” (Douglas, 2014).

Mental integrity is also protected under the Convention on the Rights of Persons with Disabilities (CRPD), an international human rights treaty of the United Nations intended to protect the rights and dignity of persons with disabilities. Unlike the CFR and the CRPD, the ECHR makes no explicit reference to the protection of physical and/or mental integrity.

Just like between physical and psychological harm, drawing a hard separation line between physical and mental integrity is a conceptually convoluted and aporetic task, as mental functions and faculties are caused and enabled by physical processes. For this reason, attempts to preserve and respect the integrity of the human being in a more holistic way (i.e., encompassing both physical and mental integrity) are more apt to overcome this dualism. For example, Article 1 of the Oviedo Convention may offer a suitable basis to such holistic approach as it states that the convention “shall protect the dignity and identity of all human beings and guarantee everyone, without discrimination, respect for their integrity” […].

The right to mental integrity is intimately intertwined with the right to life, as the protection of private life encompasses a person’s physical and psychological integrity. Mental integrity also has affinities with normative principles for the protection of people who have a mental disorder. Notably, Article 7 of the Council of Europe’s Oviedo Convention (“Protection of persons who have a mental disorder») defines the conditions under which people who have a mental disorder may or may not be subjected to an intervention without their consent. From this perspective, mental integrity can be seen in continuity with Article 7 of the Oviedo Convention and seems well-suited to also protect persons who have a mental disorder.

Regarding the risk of discrimination, while genetic discrimination is prohibited among others by the Council of Europe’s Oviedo Convention (Article 11) and the United States Genetic Information Nondiscrimination Act of 2008, no explicit safeguard against neurodiscrimination exists today. Therefore, a reconceptualization of the right should aim not only to protect against mental illness, but also to delimit the domain of legitimate manipulation of neural processing, and to prevent both brain interventions that cause mental harm and brain data processing practices that may result in neurodiscrimination.

Apart from peaceful purposes, protecting mental integrity also requires defining the limits of exploring and modulating brain activity for military usages. The laws of war that are applicable during armed conflict (so-called jus in bello or international humanitarian law) do not explicitly protect combatants against the violation of their mental integrity. There is a need to draft new normative principles similar to those guiding autonomous weapons that protect soldiers against the offensive use and misuse of neurotechnology during both wartime and peacetime. This is all the more important considering recent progress in the militarization of neurotechnology and AI, which opens the prospect of weaponizing neurotechnology and brain data (Rickli & Ienca, 2021).

3.3.4. Personal identity

In philosophy, particularly in the philosophy of mind, personal identity is typically defined as the unique identity of a person—who is usually considered subject of consciousness—over time. Personal identity is often described as a set of core properties that define someone as an individual person or make someone the person she/he is, and which distinguish them from others. These properties are typically identified subjectively (e.g., via introspection) as those properties to which the subject feels a special sense of attachment or ownership. For this reason, personal identity is usually more subjective than other forms of identity such as national identity.

The notion of personal identity often presupposes a notion of personhood, i.e., the status of being a person as opposed to a nonperson. Most philosophers interpret personhood in terms of possessing a certain set of mental properties (Baker, 2000). Accounts of personhood based on mental properties can be expressed through the sentence: ‘Necessarily, subject X is a person at time *t* if and only if properties *x, y, z* apply”. According to some authors, it may not even be necessary to have those mental properties at time *t*, as it would be sufficient to be capable of acquiring those properties (Chisholm 1976: 136f.), or to belong to a group whose members typically have those properties (Wiggins 1980: ch. 6). Although there is a general agreement that mental properties are relevant to personhood, there is ample disagreement with regard to determining which mental properties are constitutive of personal identity. Potential candidates include self-awareness, proprioception, and the capacity to suffer (Garrett, 2002; Noonan, 2019; Price-Williams, 1957).

Besides being enabled by a certain set of mental properties, personal identity is also characterized by the temporal quality of persistence, namely the quality of persisting from one time to another (Dainton & Bayne, 2005; Noonan, 2019). Most people, in fact, perceive themselves as maintaining the same individual identity over time despite their continuous replacement of bodily cells and novel life experiences. The issue of persistence of personal identity is addressed by the so-called psychological continuity theories of identity (Schechtman, 1994). These theories define personal identity in terms of overlapping chains of psychological connections that are appropriately caused. These psychological connections may involve memories or other cognitive or affective states such as, for instance, an intention and the action carried out by such intention, or between different temporal portions of a continuing belief.

One of the earliest accounts of psychological continuity can be traced back to the XVII Century philosopher John Locke, who argued that a person is “a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, *in different times and places*” (1975: 335). This implies that someone is a person if and only if they have the capacity to consider themselves in continuity with their past self, and retain the mental properties constitutive of their identity despite the passage of time and changing places. According to most accounts of psychological continuity persistence consists in some psychological relation between past, present, and future mental properties. At any given time, a person inherits mental properties such as beliefs, memories, and preferences, from their past self.

Ienca and Andorno argued that protecting this concatenation of mental properties constitutive of a person's identity is necessary to protect that person's dignity and fundamental rights (Ienca & Andorno, 2017). In particular, they argued that a right to **psychological continuity** may be necessary to prevent unwanted breaks in the chain of personhood-causing mental states due to unauthorized exogenous interventions (such as, for example, through coercive neurostimulation).

In legal theory, the right to personal identity is everyone’s right to form an individual identity, develop a conscience, and protect such individual identity and conscience from external limitation, manipulation or erasure (Marshall, 2014). According to this account, the right to personal identity is intimately intertwined with the right to life as it is only through existing that individuals can cultivate their identity. This right is recognised in international law through a range of declarations and conventions. For example, the European Court of Human Rights (ECHR) interpreted Article 8 of the European Convention on Human Rights as to include "personal identity" within the meaning of "private life", whose protection against unwanted intrusion is explicitly protected[3]. In another case ruling *(Bruggemann and Scheuten v Germany)* the ECHR highlighted the significance for personal identity of relationships concerning the "emotional field" and "the development of one’s own personality».

3.4. A CONCEPTUAL TAXONOMY OF NEURORIGHTS

While the prominence of the neurorights debate in the public opinion is crucial to ensure public engagement and democratic participation in deliberative processes on this issue, its relative sporadic nature in the academic literature poses a risk of semantic-normative ambiguity and conceptual confusion[4]. This risk is exacerbated by the presence of multiple and not always reconcilable terminologies. Most importantly, several meta-ethical, normative ethical and legal questions need to be solved. For these reasons, in this section we will try to provide a systematic classification of the

[3] See: *Goodwin v the UK* (2002) 35 EHRR 18 at 90.
[4] A keyword search of «neurorights» in the Google search engine retrieved over 22 thousand results. The same keyword search in Google Scholar retrieved slightly more than one hundred entries (stand: June 2021).

neuro-rights proposed so far. Finally, in the next section, we will discuss the main conceptual issues still open. We can identify at least five thematic families of neurorights, depending on the normative ethical principles from which they are derived: derivatives of freedom of thought, derivatives of privacy, derivatives of mental integrity, derivatives of personal identity and other ethical corollaries.

3.4.1. Thematic area: freedom of thought

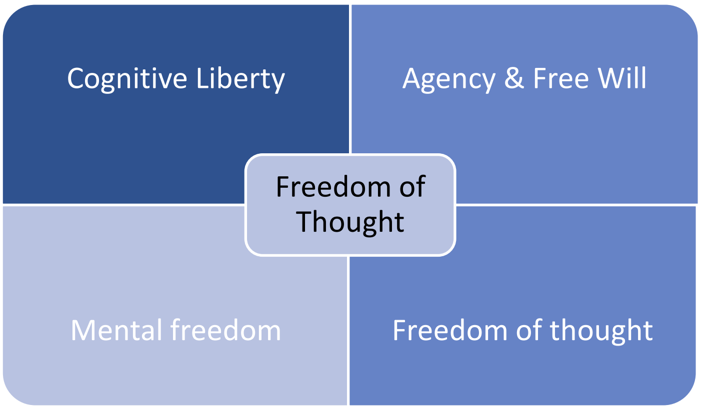

Figure 4- Freedom of thought and related concepts

In the neurorights debate, four rights conceptually derived from or associated with freedom of thought have been proposed. These are: the right to cognitive liberty, the right to agency and free will, the right to mental freedom and the right to freedom of thought itself.

As we have seen earlier, cognitive liberty was a precursor to the neuro-rights debate. Despite some minor conceptual differences in the interpretation of this right, there is a general consensus in the literature that cognitive liberty entails a person's autonomous, unimpeded control over her mind. This view is well-captured by Bublitz's use of cognitive liberty as a synonym for "mental self-determination" (Bublitz 2013). According to Bublitz (2013), this right comprises two fundamental and intimately related principles: (a) the right of individuals to use emerging neurotechnologies; (b) the protection of individuals from the coercive and unconsented use of such technologies. In other words, cognitive liberty is the principle that guarantees "the right to alter one's mental states with the help of neurotools as well as to refuse to do so" (Bublitz 2013, p. 234). Analogously, Ienca and Andorno pointed out that cognitive liberty is a "complex right which involves the prerequisites of both negative and positive liberties» in the sense of Isaiah Berlin (Berlin, 1969): the negative liberty of making choices about one's own cognitive domain in absence of external obstacles, barriers or prohibitions; the negative liberty of exercising one's own right to mental integrity in absence of external constrains or violations; and finally, the positive liberty of having the possibility of acting in such a way as to take control of one's mental life (Ienca & Andorno, 2017b). Farahany added a phenomenological layer to these definitions as she defined cognitive liberty as "the right to self-determination over our brains and mental experiences" (Farahany, 2019) (p.97). Furthermore, she provided the most detailed and comprehensive analysis to date of the range of freedoms and entitlements encompassed by this right. In her view, these include "freedom of thought and rumination, the right to self-access and self-alteration, and to consent to or refuse changes to our brains and our mental experiences" (p. 98). Finally, she emphasized that the right to cognitive liberty, like all individual interests, “is not absolute absolute, but must be balanced against the societal costs it introduces» (ivi).

While there is general agreement on the basic premises of cognitive liberty, there is disagreement with regard to its domain of application. Most definitions, including Bublitz’s definition above, limit the purview of cognitive liberty only to alterations of mental states “induced” by “neurotools” or “neurotechnologies”. In the same article, Bublitz proposes an even narrower definition of cognitive liberty which is restricted to the use of neurotechnology for the purpose of neuroenhancement (p.233). This definition, accordingly, seems to exclude alterations of mental states that do not enhance brain function (e.g., those that diminish it or cause qualitative instead of quantitative changes). In contrast, and drawing upon previous work on online manipulation (see Susser et al. 2018), Ienca and Vayena have proposed a broader and medium-independent definition which also envelops unintended alterations of mental states induced by non-neurotechnologies such as via social media and online manipulation (Ienca & Vayena, 2018), regardless of whether they result in enhancement, diminishment or non-variation of brain function. Similarly, Farahany (2019) has identified a broad range of applicability of cognitive liberty to tort law. As the kind and variety of technologies capable of decoding and/or altering the human brain is constantly evolving, technology-neutral accounts of cognitive liberty appear preferable over technology-dependent ones.

Yuste et al. have advocated a “right to agency, or the freedom of thought and free will to choose one’s own actions» Yuste et al. (Yuste et al., 2021). Although these authors use those three notions as synonyms, as indicated by the disjunctive logical operator “or”, agency, freedom of thought and free typically denote quite distinct concepts. Agency, as it is widely discussed in the philosophy of action literature, denotes the exercise or manifestation of an agent’s capacity to act (Gallagher, 2007; Proust, 2013). Free will, as we have seen, is an ontological thesis related to the capacity of agents to choose between different courses of action unimpeded (Dennett, 2015; Spence, 1996). In other words, agency pertains to the domain of action. Free will, in contrast, pertains to the domain of cognition, decision-making in particular. Most importantly, both agency and free will are typically conceptualized as abilities or dispositions. They are descriptive in nature, not normative (Aarts & van den Bos, 2011). Deriving normativity from these descriptive statements requires inferring entitlements and obligations from abilities and dispositions. The logic of such inference, however, remains currently unclear. Finally, as observed by Munoz, “free will is a multidimensional concept that poses several unsolved philosophical problems» (Munoz, 2019).

Mental freedom is used seldom in the literature. Repetti used mental freedom to outline his Buddhist theory of free will (Repetti, 2018). A specific use of mental freedom (which he also calls “freedom of mind”) in the context of neurorights is provided by Bublitz who described it as the “conscious control over one’s mind”(C. Bublitz, 2016).

He argued that mental freedom should be ranked among the most important legal and political freedoms (ivi). It is unclear, however, if his notion of mental freedom (which he also calls “freedom of mind”) is equitable to cognitive liberty or rather conceptually distinct from it.

Finally, some authors have argued that the very notion of freedom of thought offers suitable normative ground to address the human rights challenges raised by novel neurotechnologies (Lavazza, 2018). Adopting freedom of thought as the normative foundation of a person’s autonomous control over her mind is advantageous from a conceptual parsimony perspective. The Occam’s razor principle or law of parsimony postulates that "entities should not be multiplied without necessity"(Schaffer, 2015). Since freedom of thought is already enshrined in international human rights law and widely discussed in legal philosophy, it would be ceteris paribus more parsimonious to adopt this normative terminology compared to multiplying the number of normative entities by introducing cognitive liberty, mental freedom and the rights to agency and free will. Should this parsimonious approach be pursued, however, it should be clarified that “the protection of a person’s self-determination over her mind should comprise the entire forum internum» (C. Bublitz, 2015), that is all mental states or capacities and there-with cognitive, emotional and conative phenomena, either conscious or unconscious”. As Ienca and Andorno pointed out, freedom of thought is the fundamental justification of related freedoms such as freedom of choice, freedom of speech, freedom of press, and freedom of religion. An evolutionary interpretation of this right should focus on the protection not only of the externalizations of thought but thought itself.

A possible objection against this broad conceptualization, however, is that the notion of freedom of thought, as we have observed earlier and attested by the UDHR and the ECHR, is historically intertwined with the notions of freedom of conscience and religion. For this reason, its original purpose and historical function may be irreconcilable with the contemporary need to protect the *forum internum* and to guarantee the freedom of persons to self-determine in relation to their own minds. For this reason, some authors have argued that freedom of thought and cognitive liberty are not simply different nomenclatures but conceptually distinct rights. For example, Farahany has argued that cognitive liberty is broader than freedom of thought as it encompasses, besides the freedom of thought, also the rights to self-access and self-alteration, and the right to consent to or refuse changes to our brains and our mental experiences (Farahany, 2019). From this point of view, albeit conceptually parsimonious, assimilating cognitive liberty to freedom of thought would be unjustified. Consequently, the introduction of new rights such as the right to cognitive liberty would not represent a mere unnecessary multiplication of normative entities, but a necessary clarification of the human rights framework in the light of technological evolution.

3.4.2. Thematic area: privacy

Unlike the derivatives of freedom of thought, the neurorights originating from the right to privacy seem to be characterized by a much greater degree of conceptual and terminological agreement. **Mental privacy** is the expression generally used to denote people's right against the unconsented intrusion by third parties into their mental data as well as against the unauthorized collection and processing of those data (Ienca & Andorno, 2017a, 2017b; Shen, 2013; Yuste et al., 2021). Yuste et al. argued that mental privacy is not only a right but also an ability, i.e., «the ability to keep thoughts protected against disclosure» (Yuste et al., 2021). The relationship between mental privacy and the general right to private life is debatable. Ienca and Andorno argued that the special nature of brain information requires attaching additional specifications to the current privacy frameworks. This is, in their view, due to the fact that brain data relate directly to one's inner mental life and personhood and are obtained in a distinctive way. They argued that mental privacy should protect brainwaves not only as data but also as data generators or sources of information. They called this challenge the "inception problem". In addition, a right to mental privacy would protect not only conscious brain data but also data that are not (or are only partly) under voluntary and conscious control. Finally, it would guarantee the protection of brain information in absence of an external tool for identifying and filtering that information. This would contribute to protecting people's right against illegitimate access to their brain information and to preventing the indiscriminate leakage of brain data across the infosphere.

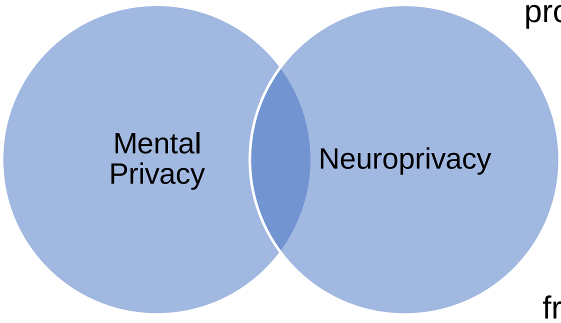

Figure 5- Mental Privacy & Neuroprivacy

Authors have also debated how to facilitate the applicability of the right to mental privacy in binding regulation. For example, Ienca & Malgieri have proposed introducing a Mental Data Protection Impact Assessment (MDPIA), namely a data protection impact assessment involving an audit of the neurotechnological components of the data processing (e.g., the AI algorithms) and a reconsideration of the algorithm in case some risks can be mitigated "by design" (Ienca & Malgieri, 2021). In their view, data controllers of mental data should be obliged to comply with the following requirements:

a) Describing the processing (including a description of the logic of the technology)[5]
b) Performing a balancing test based on necessity and proportionality of the data processing in relation to the purposes,[6]
c) Assessing the actual risks for fundamental rights and freedoms and proposing suitable measures to address and mitigate those risks.

Another concept frequently used to address people's moral entitlement to protect their brain information is neuroprivacy. While 'mental privacy' aims at protecting mental information, however collected or inferred, neuroprivacy relates specifically to the protection of neural data—also called neurodata or brain data (Hallinan, Schütz, Friedewald, & de Hert, 2014; Ienca, 2015; Wolpe, 2017). The difference between these two notions is not trivial. Mental information is information about mental states such as thoughts, memories, beliefs, perceptions and emotions. Acquiring mental information requires some degree of mental decoding. Neural information, in contrast, is information about the nervous system. Although these two types of information have a significant area of overlap, they are not supervenient. It is possible to access neural information without accessing mental information and vice versa. For example, by gaining access to a person's neuroimaging biomarkers of Alzheimer's disease (e.g., a measurement of beta-amyloid deposition with amyloid PET), I thereby gain access to that person's neural information. However, I do not automatically gain any access to that person's mental states. Vice versa, if non-neural technologies such as face recognition and emotion AI are used to reveal information about a person's emotional state (e.g., a feeling of fear), they thereby ensure a certain degree of access to mental information. However, they do not automatically provide any access to that person's neural information. Therefore, mental privacy and neuroprivacy appear to be complementary, non-reducible instantiations of privacy in the bio-digital world.

3.4.3. Thematic area: mental integrity

A relatively strong conceptual convergence is also recognizable with regard to mental integrity. As we have seen before, the right to mental integrity is enshrined in the EU's Charter of fundamental rights (Article 3). However, divergences exist with regard to how this right is interpreted. Ienca and Andorno argued that the right to mental integrity might require an evolutionary interpretation as to better safeguard people's right to be protected from illicit and harmful manipulations of their mental activity. In contrast, Douglas defined 'mental integrity' as "a right against mental interference" (Douglas, 2014). Similarly to Douglas, Lavazza interpreted mental integrity as "the individual's

[5] Margot E Kaminski and Gianclaudio Malgieri, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' [2020] International Data Privacy Law <https://doi.org/10.1093/idpl/ipaa020> accessed 1 February 2021.
[6] Dariusz Kloza and others, 'Data Protection Impact Assessment in the European Union: Developing a Template for a Report from the Assessment Process' (LawArXiv 2020) DPiaLab Policy Brief <https://osf.io/7qrfp> accessed 1 December 2020.

mastery of his mental states and his brain data so that, without his consent, no one can read, spread, or alter such states and data in order to condition the individual in any way» (Lavazza, 2018). The conceptual difference here is important. While Douglas and Lavazza interpret mental integrity in close proximity with the right to cognitive liberty and the right to freedom of thought, Ienca's and Andorno's definition establishes a necessary logical relationship between mental integrity and the protection from harm related to someone's neural and/or mental domain. In the first case, it would follow that mental integrity is substitutive of or subsumed by cognitive liberty and freedom of thought. In the second case, it is complementary to them.

Douglas' work on medical interventions for criminal rehabilitation can help us clarify this point. In his view, the nonconsensual imposition of medical correctives that "are intended to have mental effects" might be thought to constitute a violation of the right to mental integrity because it is a kind of nonconsensual mental interference, hence "a serious threat to agency" (Douglas, 2014). According to Ienca & Andorno's definition, this type of medical correctives always constitutes a violation of the recipient's cognitive liberty. However, they constitute a violation of mental integrity only if such correctives cause harm to the subject.

In addition to these conceptual considerations, enshrining the right to mental integrity requires clarification of the relationship between that right and the right to bodily integrity. As stated earlier, drawing a hard separation line between physical and mental integrity is an aporetic task, which is rooted in a (either implicitly or explicitly) dualist ontology of the mind-body relation. Since mental functions and faculties are caused and enabled by physical processes, holistic attempts to preserve and respect for the integrity of the human being are better suited to overcome this dualism. Therefore, neurorights accounts of mental integrity should avoid dualistic premises and provide a comprehensive framework for the protection of personal integrity which includes both the human body and the faculties and processes enabled by the body.

A good example of this holistic approach, as mentioned earlier, is Article 1 of the Oviedo Convention. Similarly, the Irish Supreme Court pronounced, in 1965, that "you have the right not to have your body or personhood interfered with" (Ryan v Attorney General). According to this pronouncement, physical and mental integrity are to be seen as one single and necessary prerequisite for personal identity. In a similar fashion, the Fourth Amendment to the United States Constitution entails the protection of "the right of the people to be secure in their persons», suggesting a broader view of integrity which envelops both physical and mental integrity.

Drawing on comments previously made by others to motivate the recognition of this right, Douglas & Forsberg (2021) have identified and outlined three distinct rationales for recognizing a legal right to mental integrity: the appeal to intuition, the appeal to justificatory consistency, and the appeal to technological development. The appeal to intuition regards the moral intuition that people should be protected from mental interferences, especially those "causing mental suffering" (Bublitz & Merkel, 2014). The appeal to justificatory consistency is the view that standard theoretical justifications for the right to bodily integrity appear also to support a right to mental integrity. The appeal to technological development is the view that the disruptive nature of recent and likely future neurotechnological developments requires heightened protection of the mental sphere. Douglas & Forsberg conclude that each of these rationales is open to question. Nevertheless, they argue that «each of these candidate rationales has some plausibility and warrants further scrutiny» (Douglas & Forsberg, 2021).

3.4.4. Thematic area: personal identity and psychological continuity

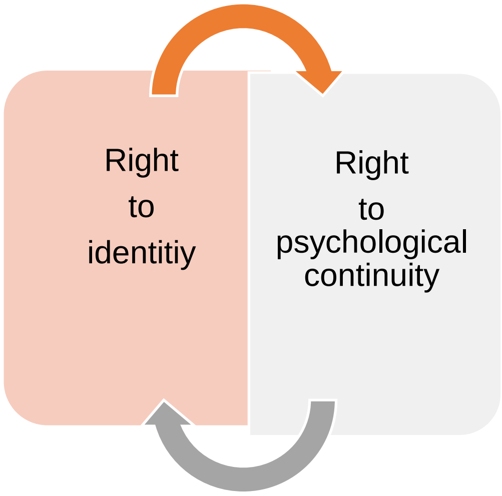

Figure 6- Identitiy and Psychological Continuity

Some authors have advocated the recognition of a fourth family of neurorights related to the protection of personal identity. Borrowing the terminology from the psychological-continuity account of personal identity (Van Inwagen, 1997), Ienca and Andorno called this right "psychological continuity" and described it as the right to preserve «people's personal identity and the continuity of their mental life from unconsented external alteration by third parties» (Ienca & Andorno, 2017a). Yuste et al., in contrast, advocated a «right to identity», which they described as «the ability to control both one's physical and mental integrity"(Yuste et al., 2021). While psychological continuity, in its original formulation, has thematic affinities to cognitive liberty and freedom of thought, the right to identity in Yuste et al.'s sense appears to be a prerequisite for physical and mental integrity. Therefore, these two rights seem to protect different, albeit complementary, human interests.

Both the right to identity and the right to psychological continuity offer solid normative ground to guide the responsible integration of AI into the control of BCI and preserve a person's self-determination and sense of personal identity from subconscious manipulation. These rights can help BCI users retain control over their behaviour, without experiencing 'feelings of loss of control' or even a 'breakdown' of personal identity (Gilbert et al., 2017). Along these lines, Clausen et al. have argued that "veto controls" should be included in BCIs to protect the autonomy and identity of users. At the same time, the right to psychological continuity is intended to protect against unauthorised interventions by third parties into the neural activity of BMI users. For instance, such a right could protect individuals from unauthorised neuromodulation, even if such interventions do not cause injury or trauma to the user (which would be protected by the right to mental integrity) but still generate significant changes in the individual's psychological sphere (e.g., in terms of political and/or religious preferences, emotional states and memories) without the latter's explicit consent. This principle may become particularly important in the context of national security and military research, where neurotechnology applications that modulate personality traits (e.g., neurostimulation techniques) are currently being tested for combatant enhancement and other strategic purposes, e.g., to increase the ability of soldiers and other military personnel to perform with motivation and determination even under stress or in the absence of sleep (Tennison & Moreno, 2012). Such neuromodulation technologies are often implemented using hybrid devices that can read signs of decreased attention through EEG and then modify these processes through stimulation.

3.4.5. Other ethical corollaries

Finally, some authors have proposed the recognition of rights that are not directly related to the protection of the neural and mental domain but rather to the promotion of the socio-technical requirements that are instrumentally necessary for the realization of the rights above. Two of these normative corollaries have been proposed: the right to fair access to mental augmentation and the right to protection from algorithmic bias. The former is defined by Yuste et al. as “the ability to ensure that the benefits of improvements to sensory and mental capacity through neurotechnology are distributed justly in the population» (p.160-161); the latter is defined by the same authors as “the ability to ensure that technologies do not insert prejudices» (ivi). As such, the right to fair access to mental augmentation appears to be a prerequisite for cognitive liberty in the positive sense. This principle can be seen in close normative proximity with Article 3 of the Oviedo Convention, which requires and promotes equitable access to healthcare of appropriate quality. It is questionable, however, if cognitive enhancement falls under the scope of healthcare. In contrast, the right to protection from algorithmic bias appears to be a prerequisite for the right to mental integrity as it protects from the spectrum of harms generated by algorithmic bias, first and foremost algorithmic neurodiscrimination. It is worth noting that unlike all other neuroright candidates above, the right to protection from algorithmic bias can be and has been advocated in domains that unrelated to the mental and/or neurocognitive sphere (Garcia, 2016). Further, it could be argued that neurodiscrimination should all ways be prohibited, regardless of whether such discrimination is caused by an algorithm or a human decision-maker. For example, the Oviedo Convention prohibits “all forms of discrimination based on the grounds of a person's genetic make-up”. Analogously, neurorights should prohibit all forms of

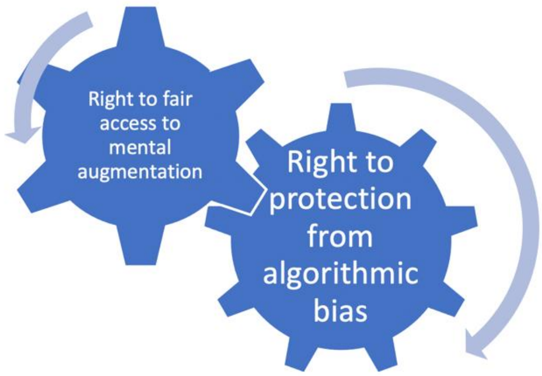

Figure 7- Other ethical corollaries

discrimination based on the grounds of a person's neural make-up and/or mental states.

Finally, it is surprising to note that positive neurorights such as the right of people to access and use neurotechnologies have only rarely been explored. For example, the UN Convention on the Rights of Person with Disabilities (CRPD) obliges states to ensure and promote the capacity of people with disabilities to live independently and participate in social life. This includes efforts in research and development of assistive devices, and affording patients access to them (e.g. art. 4g CRPD). Future research should determine if such right also extends to neurotechnology use among healthy individuals for purposes such as cognitive enhancement, as a positive characterization of the right to cognitive liberty and the the right to fair access to mental augmentation would entail.

Most importantly, further research is needed to explore positive neurorights such as promoting patient welfare on the basis of the ethical principle of beneficence, which have so far occupied a secondary role in the neurorights debate.

4. NEUROIGHTS IN INTERNATIONAL POLICY

In the space of a few years, neurorights have moved from the mere realm of theoretical reflection to becoming the subject of debate in international politics. To date, several governmental, intergovernmental, and non-governmental actors who are actively involved in neurotechnology governance have included or are including neurorights on their agenda.

An important first step in introducing neurorights as regulatory tools was marked in 2019, when the Council of the Organization of Economic Development and Cooperation (OECD) adopted a *"Recommendation on Responsible Innovation in Neurotechnology"*(OECD-Council, 2019). The OECD Recommendation is a soft-law instrument designed to anticipate and address the ethical, legal and social challenges raised by novel neurotechnologies while promoting innovation in the field. As such, it set the first internationally accepted standard in neurotechnology governance. The OECD Recommendation is primarily focused on responsible governance by neurotechnology industry actors. Nonetheless, it features provisions on three classes of neurorights: mental privacy (with an emphasis on safeguarding brain data and other information), mental integrity (with an emphasis on anticipating and monitoring potential unintended use and/or misuse), and cognitive liberty.

The OECD is not the only intergovernmental organization that is putting neurorights at core of their governance strategies. In 2020, the Council of Europe has launched a five-year Strategic Action Plan focused on Human Rights and Technologies in Biomedicine, which contains a module on the assessment of the relevance and sufficiency of the existing human rights framework to address the issues raised by the applications of neurotechnologies. The aim of this program is to assess whether the fundamental ethical-legal issues raised by neurotechnology "can be sufficiently addressed by the existing human rights framework or whether new human rights pertaining to cognitive liberty, mental privacy, and mental integrity and psychological continuity, need to be entertained in order to govern neurotechnologies». This strategic action plan is in continuity with the regulatory framework introduced by the Oviedo Convention, namely preventing the misuse of innovations in biomedicine and protecting human dignity.

In parallel, the Committee on Legal Affairs and Human Rights (Rapporteur: Mr Olivier BECHT, France) has produced a report entitled “The brain-computer interface: new rights or new threats to fundamental freedoms?” (Report Doc. 15147 | 24 September 2020). The committee argued that a calibrated approach to the regulation of BCI technology is needed, encompassing both ethical frameworks and binding legal regulation. It therefore called on member States, the relevant intergovernmental committees of the Council of Europe and the Committee of Ministers to take specific steps to this end.

National legislators are also active in the area of neurotechnology governance. The most important policy development in this area is the recent approval by the Chilean Senate of a constitutional reform law that defines mental integrity as a fundamental human right, and a law on neuroprotection that protects brain data and applies existing medical ethics, codified in the current Chilean medical code, to the use of neurotechnologies in non-patient populations. In April 2021, Chile's Senate Future Commission approved the final text amending article 19 of the Constitution endorsing the “rights to physical and mental integrity” and protecting “cerebral activity and its data”. This makes Chile, “the only country with a proposed law and constitutional amendment mandating neuroprotection and explicitly protecting neurorights» (Yuste et al., 2021). Furthermore, the Spanish Secretary of State for AI has recently published a Charter of Digital Rights that incorporates neurorights as part of citizens' rights for the new digital era. Finally, the Italian Data Protection Authority has devoted the 2021 Privacy Day to the investigation of neurorights and endorsed their necessity, especially mental privacy, and mental integrity, to properly address the implications of neurotechnology for human rights with special regard to the right to privacy.

5. OPEN QUESTIONS AND THE FUTURE OF NEURORIGHTS

Although (or perhaps precisely because) neurorights have moved in a relatively short time from the domain of ethical-legal reflection to that of advocacy and policy, many conceptual and practical questions remain unanswered. The first question is determining whether neurorights should interpreted as rights in the philosophical sense (also called *moral rights*), as legal rights, as human rights in the sense of international human rights law or all of the above.

Notably, Capron (1976) observed that, although intimately related, moral and legal rights are not identical as there may be circumstances under which a person has a moral right "to be able to compel another's behavior (or whatever) but legally cannot" (Capron, 1976). More radically, Thomson has argued that legal and moral rights are two discrete territories within the realm of rights because they derive from different sources: while moral rights are created by practical reason (phronesis, in Aristotelian terms), legal rights are created only by a legal system and through the process of law-making (Thomson, 1990). The overview contained in this report indicates that neurorights should be seen as both rights in the philosophical sense, i.e., as moral rights, and as legal rights, i.e., rights which exist under the rules of legal systems or by virtue of decisions of suitably authoritative bodies within them. This is consistent with John Stuart Mill's thesis of a close analytical connection between moral and legal rights, being all rights related to fundamentals of wellbeing. As previous legal theory has observed, basing legal rights on moral ones can provide a better justification in practical reasoning and policy (Hohfeld, 1913). Accordingly, any future international declaration or legislative reform related to neurotechnology and human rights could benefit from being based on neurorights as moral rights.

A second pressing question is to determine whether neurorights in the sense of international human rights law are to be interpreted as brand-new human rights or as evolutionary interpretations of existing rights. Two problem-solving principles may offer guidance in this regard. First, as we have seen, Occam's razor or law of parsimony requires that entities should not be multiplied without necessity. Second, the principle of avoiding 'rights inflation', i.e., the objectionable tendency to label everything that is morally desirable as a 'human right', postulates that the unjustified proliferation of new rights should be avoided. The unwarranted proliferation of human rights is problematic because it spreads scepticism about all human rights, as it dilutes them to mere desiderata or purely rhetorical claims. In other words, rights inflation is to be avoided because it dilutes the core idea of human rights and distracts from the central goal of human rights instruments, which is to protect a set of truly fundamental human interests, not everything that would be desirable or advantageous in an ideal world (Ienca, 2021).

From this perspective, the most parsimonious approach would be to treat neurorights as evolving interpretations of existing rights by default, while at the same time imposing justificatory tests to assess whether they actually constitute new human rights. Several justificatory tests to prevent rights inflation have been proposed. For example, Alston proposed a list of criteria that a normative claim must satisfy in order to qualify as a 'human right'. In his view, the new human right candidate must (i) "reflect a fundamentally important social value"; (ii) "be consistent, but not merely repetitive, of the existing body of international human rights law"; (iii) "be capable of achieving a very high degree of international consensus", and (iv) "be sufficiently precise as to give rise to identifiable rights and obligations" (Alston, 1984). Similarly, Nickel has required that a proposed human right should not only (i) deal with some very important good but also (ii) respond to a common and serious threat to that good, (iii) impose burdens on the addressees that are justifiable and no larger than necessary, and (iv) be feasible in most of the world's countries (Nickel, Pogge, Smith, & Wenar, 2013). In the light of the analysis of the ethical and human rights challenges raised by neurotechnology, it appears clear that the protection of a person's cerebral and mental space reflects a fundamentally important social value, aims to safeguard a very important good, and responds to a serious threat to that good.

Most of the neurorights proposed so far, appear to impose burdens that are justifiable and no larger than necessary, as none of the normative reviewed in this report appears to temper neurotechnology innovation or exert a disproportionate burden on neurotechnology developers, researchers and other stakeholders. Additionally, neurorights appear consistent with the existing body of international human rights law, as they complement fundamental rights and freedoms; but, at the same time, they introduce normative specifications related to the protection of the person's mental domain that are not merely repetitive of that framework. Although neurotechnologies are unevenly distributed on a global scale, the protection of neurorights concerns all world's countries. However, it is too early to determine whether and which neurorights will be capable of achieving a high degree of international consensus. The recent approval by the Chilean Senate of a neuroprotection bill which promotes the rights to mental privacy and mental integrity, and the relative convergence around these two rights in the ethical-legal scholarship, makes these two rights strong candidates for inclusion in the international human rights framework. Finally, further research is needed to assess whether and which neurorights are sufficiently precise as to give rise to identifiable rights and obligations.

As the analysis above attests, for the field of neurorights to advance and have consistent impact on international policy, it should overcome the current terminological variations and semantic ambiguities related to how these neurorights are denominated, defined, and interpreted. Without a common terminology, semantic disambiguation, and conceptual harmonization, it is unlikely that neurorights-based initiatives will lead to effective national and international policies. This process of harmonization should not obliterate divergent views. On the contrary, it should include them in a pluralistic and deliberative democratic manner. Furthermore, it should ensure that neuroright proposals are adequately vetted, conceptually demarcated, normatively justified and rooted in both moral philosophy and existing regulations.

This analysis suggests that the right to mental privacy has the largest consensus and strongest conceptual stability in the international debate. This is due to a general agreement on the need to protect brain and mental data from unauthorised access, inspection, and modification. With regard to the right to mental integrity, the debate is open as to whether protection from harm is a constitutive requirement of this right. However, there is broad consensus on the need to protect the person from mental interferences. Finally, the neurorights candidates belonging to the thematic areas of mental self-determination (i.e., freedom of thought, cognitive liberty, mental freedom) and personal identity (i.e., rights to personal identity and psychological continuity) are characterized by a higher semantic and interpretative variation. However, despite multiple, often irreconcilable terminologies, there is a general agreement that the freedom to exercise control over one's cognitive and emotional sphere is the prerequisite for all other freedoms and an essential prerequisite for the principle of personal autonomy and self-determination.

A third question regards how neurorights can be adequately implemented and enforced. Should in the future some of the neurorights described in this paper pass several justificatory tests and obtain strong democratic and deliberative support, how should they be enforced? There are two types of human rights instruments: declarations and conventions. Declarations are not legally binding but do have political impact, whereas conventions are legally binding under international law. Both declarations and conventions can become customary international law over time, which makes them universally legally binding (Moscrop, 2014). The 1997 UNESCO Universal Declaration on the Human Genome and Human Rights and the 2005 UNESCO Declaration on Bioethics and Human Rights may offer suitable ground and reference to explore the development of an analogous declaration focused on the human brain and human rights. Among the conventions, the Council of Europe's

“Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine” (Oviedo Convention) provides the most suitable and comprehensive model for future initiatives and international instruments aiming to protect the human brain. Given its focus on prohibiting the misuse of innovations in biomedicine, protecting the dignity and identity of all human beings, and guaranteeing respect for their integrity and fundamental freedoms, the Convention is perfectly placed for enshrining neurorights such as the rights to mental integrity, personal identity and cognitive liberty. Neurorights could be included as an additional subject-specific protocol which further elaborates the provisions of the Convention.

Future legal scholarship should discuss which type of instrument is most suitable for enshrining neurorights into international human rights law. Further, it should determine how the problem of “under-enforcement” of human rights can be avoided (Koh, 1998), that is how to achieve state obedience of neurorights laws from a realist perspective.

Finally, future scholarship should discuss the place of neurorights within neurotechnology governance. Unless one commits to the unlikely thesis that neurorights are sufficient for neurotechnology governance (hence that neurotechnology governance can be entirely reduced to neurorights promotion), it is critical to clarify how neurorights relate to other governance mechanisms such as self-regulation by neurotechnology actors, ethical guidelines, and binding regulations in areas such as inter alia health law, data protection law, consumer protection law and criminal law.

6. CONCLUSIONS AND POLICY OPTIONS

Since the dawn of mankind, technological innovation has been a major catalyser of societal and cultural transformation. Our ancestors have constantly modified their surroundings through the development of artifacts whose complexity has gradually increased through cultural evolution. The *homo sapiens* (literally “man the wise”) is fundamentally a *homo technologicus* (i.e., “man the user of technology”). These increasingly complex technologies have co-evolved with humans and human societies. In some cases, major socio-technological transformations such as the invention of agriculture, writing and modern medicine have exerted a profound impact on human life and the very meaning of being human. Throughout this process of technological evolution, human societies have introduced systems of norms, be it socially accepted customs, church-mandated recommendations or binding laws enforced by the state.

Among these systems of norms, human rights constitute the most fundamental and universal categories. This is because human rights are inherent to all human beings regardless of nationality, sex, national or ethnic origin, color, religion, language, or any other status. As human rights are rights that ought to be respected simply because we exist as human beings, their conceptual definition and articulation are interdependent with what it means to be a human. If the meaning of being human evolves as a consequence of socio-technological change, then human rights may evolve accordingly. This evolution is necessary to protect those fundamental freedoms and entitlements that make human life worth living at any given historical time.

The European Convention on Human Rights is a paradigm example of such normative evolution. In over sixty years since its adoption, the ECHR has constantly evolved as a consequence of legal-philosophical reflection and the European Court of Human Rights’ case law. For this reason, the Convention has been often regarded as a “living instrument", which is able to adapt to the changes taking place in our societies. This process of constant adaption has also involved the addition of new rights to the Convention, whenever fundamental entitlements emerged with regard to situations that could not have been anticipated when the Convention was first adopted.

In the last century, innovation in the areas of both biomedical and information and computing technology has generated highly powerful and transformative technologies. These technologies hold unprecedented potential for impinging on human rights. Among these technologies, neurotechnology and Artificial Intelligence (and the combination thereof) play a critical role as they generate novel opportunities for monitoring, influencing, altering and simulating the human brain and mind. Being the human brain and mind the physical and functional substrates of fundamental human faculties such as personal identity, consciousness language and emotions, all technologies that can monitor, influence, alter and simulate the human brain and mind have an impact for human rights. As a consequence, the evolution of these technologies urges our societies to reflect on those impacts and determine whether a normative co-evolution may be necessary to protect and promote human rights in the light of these technologies.

It is the task of empirical science, especially neuroscience, psychology, anthropology and medicine, to study how these technologies may change humans in the years to come. As these technologies open the prospect of intervening into human capabilities at the neurobiological and information processing level, it plausible to predict that the current neurotechnological revolution will be at least as transformative of human nature as previous epoque-changing inventions.

At the same time, it is the imperative task of normative disciplines such as ethics and the law, to determine which aspects of such technology-induced anthropological transformation are morally desirable and legally acceptable. In particular, we ought to determine which rights people are entitled to have about the brain and mind. As neurotechnology holds disruptive potential for changing salient features (e.g., memory, consciousness, reasoning, emotions, beliefs and desires) of the human, it holds thereby a potential for changing the fundamental rights, freedoms and obligations of the human (Ienca, 2019b).

This report attempted to provide a detailed overview of the major technological developments in the areas of neurotechnology and AI, with special focus on brain-computer interfaces. Further, it attempted to summarize the theoretical reflection on the fundamental moral and legal entitlements relating to the use of these technologies. Finally, it provided a descriptive account and conceptual taxonomy of neurorights, i.e., the moral, legal, social or natural principles of freedom or entitlement related to a person's brain and mind.

This analysis suggests that neurorights reflect fundamental human interests that are deeply rooted in the history of ideas. These rights introduce normative specifications related to the protection of the person's mental and neural domain that are not merely repetitive of existing human rights frameworks. Further, it corroborates the view that the fundamental rights and freedoms relating to the human mind and brain are the fundamental substrate of all other rights and freedoms. Therefore, protecting neurorights is a fundamental task of international human rights law.

Our overview indicates that there is not yet complete consensus regarding the conceptual-normative boundaries and terminology of neurorights. Divergences exist in relation to how these rights are interpreted, worded, and conceptually articulated. Nonetheless, some degree of convergence is emerging around three families of neurorights. First and foremost, the need for specific provisions on the protection of private brain-related information (mental privacy and neuroprivacy) seems to share a high degree of acceptance and recognition. Second, the right to mental integrity appears to have the have the highest degree of legal entrenchment. While there are some variations in interpretation, there is full theoretical consensus about the need to protect the person from psychological harm and mental interference. Third, a variety of neurorights candidates have been proposed to preserve and promote the freedom of the human mind and thereby prevent external manipulation. These include evolutionary interpretations of the right to freedom of thought, the right to cognitive liberty, and the right to personal identity.

Surprisingly, positive rights such as promoting justice and equality, e.g., through ensuring egalitarian access to neurotechnology for biomedical use and promoting patient welfare on the basis of the ethical principle of beneficence, have so far occupied a secondary role in the neurorights debate.

The above-listed families of neurorights appear deeply rooted in the international human rights framework and legal doctrine. However, they are insufficiently specified in current human rights instruments such as the UDHR, the ECHR and the CFR. Therefore, a process of normative reform appears to be needed to adequately specify principles of freedom or entitlement related to a person's mind and brain domain in the neurotechnological era. This process may occur in a twofold manner: through adaptive interpretation of existing rights and through addition of new rights.

Normative evolution in the light of disruptive technological innovation is not a new phenomenon. For example, the technological development of the mechanical ventilator produced the concept of brain death and required the law to specify more clearly what functions are integral to life and what are not, as well as how to resolve disagreements over the moral status of human beings who could alternatively be viewed as severely brain damaged but not dead (Machado, 2007). Most importantly, the ECHR has evolved as a "living instrument" as the adoption of protocols has added new rights to the Convention. These include the right to peaceful enjoyment of property, the right to education and the right to free elections by secret ballot (Protocol No. 1) and the general prohibition of discrimination (Protocol No. 12). More closely related to the biomedical technology domain, the 1997 Universal Declaration on the Human Genome and Human Rights (UDHGHR) and the 2003 by the International Declaration on Human Genetic Data (IDHGD). These instruments were both developed in response to advances in genetics and introduced new rights such as the 'right not to know one's genetic information' (UDHGHR Art. 5(c); IDHGD (Art. 10)). Similarly, the Oviedo Convention has introduced additional protocols such as a protocol on the prohibition of human cloning in response to reproductive artificial cloning techniques involving the reproduction of human cells and tissues.

It is desirable that neurorights shall follow a similar historical trajectory in a manner that expands and enhances the capacity of our human rights framework to address the profound implications of neurotechnology and AI for human nature, human dignity and human rights. Introducing neurorights into the human rights framework may require adding new protocols to existing instruments or even stipulating new multilateral instruments entirely devoted to neuroethics and neurolaw. In either case, some fundamental ethical, meta-ethical, and legal issues must be addressed in order to overcome problems such as rights inflation and to provide an adequate normative justification for neurorights. These include introducing justificatory tests for the introduction of neurorights, clarifying the relationship between moral and legal neurorights and harmonizing neurorights with existing normative instruments.

The Council of Europe's Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (Oviedo Convention) offers an ideal platform and normative substrate for the protection and promotion of neurorights. Given its focus on prohibiting the misuse of innovations in biomedicine, protecting the dignity and identity of all human beings, and guaranteeing respect for their integrity and fundamental freedoms, the Convention is well placed for either enshrining neurorights through ad hoc protocols or for serving as a basis for future instruments.

Understanding, treating, and augmenting the human brain and mind is one of the great scientific challenges of our age. Achieving these goals in a way that preserves justice, safeguards fundamental rights and human dignity is the corresponding task of ethics and law. Neurorights will likely be a useful tool to accomplish this task.

REFERENCES

Aarts, H., & van den Bos, K. (2011). On the foundations of beliefs in free will: Intentional binding and unconscious priming in self-agency. *Psychological science, 22*(4), 532-537.

Ajrawi, S., Rao, R., & Sarkar, M. (2021). Cybersecurity in brain-computer interfaces: RFID-based design-theoretical framework. Informatics in Medicine Unlocked, 22, 100489.

Alston, P. (1984). Conjuring up new human rights: A proposal for quality control. *Am. J. Int'l L., 78*, 607.

Ayaz, H., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2019). Chapter 3 - The Use of Functional Near-Infrared Spectroscopy in Neuroergonomics. In H. Ayaz & F. Dehais (Eds.), *Neuroergonomics* (pp. 17-25): Academic Press.

Baker, L. R. (2000). *Persons and bodies: A constitution view*: Cambridge University Press.

Beitz, C. R. (2011). *The idea of human rights*: Oxford University Press.

Berlin, I. (1969). Two concepts of liberty. *Berlin, i, 118*(1969), 172.

Bles, M., & Haynes, J.-D. (2008). Detecting concealed information using brain-imaging technology. *Neurocase, 14*(1), 82-92.

Bode, S., He, A. H., Soon, C. S., Trampel, R., Turner, R., & Haynes, J.-D. (2011). Tracking the unconscious generation of free decisions using uitra-high field fMRI. *Plos one, 6*(6), e21612.

Boire, R. G. (2001). On cognitive liberty. *The Journal of Cognitive Liberties, 2*(1), 7-22.

Bostrom, N., & Sandberg, A. (2009). Cognitive enhancement: methods, ethics, regulatory challenges. Science and engineering ethics, 15(3), 311-341.

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S. S., . . . Brookes, M. J. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature, 555*(7698), 657-661. doi:10.1038/nature26147

Boyden, E. S. (2015). Optogenetics and the future of neuroscience. *Nature neuroscience, 18*(9), 1200-1201. doi:10.1038/nn.4094

Brandeis, L., & Warren, S. (1890). The right to privacy. *Harvard law review, 4*(5), 193-220.

Brimacombe, P. (2000). All the Queen's Men: The World of Elizabeth I. New York: St. In: Martin's Press.
Bublitz, C. (2015). Cognitive Liberty or the International Human Right to Freedom of Thought. In J. Clausen & N. Levy (Eds.), *Handbook of Neuroethics* (pp. 1309-1333). Dordrecht: Springer Netherlands.
Bublitz, C. (2016). Moral Enhancement and Mental Freedom. *Journal of Applied Philosophy, 33*(1), 88-106. doi:https://doi.org/10.1111/japp.12108
Bublitz, J.-C. (2013). My mind is mine!? Cognitive liberty as a legal concept. In *Cognitive enhancement* (pp. 233-264): Springer.
Bublitz, J. C., & Merkel, R. (2014). Crimes against minds: On mental manipulations, harms and a human right to mental self-determination. Criminal Law and Philosophy, 8(1), 51–77. https://doi.org/10.1007/s11572- 012-9172-y.
Bury, J. B. (1914). *A history of freedom of thought*: Gorgias Press.
Capron, A. M. (1976). Legal rights and moral rights. In *Biomedical Ethics and the Law* (pp. 375-397): Springer.
Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A., & Yuste, R. (2019). Controlling Visually Guided Behavior by Holographic Recalling of Cortical Ensembles. *Cell, 178*(2), 447-457 e445. doi:10.1016/j.cell.2019.05.045
Cascio, J. (2017). Do brains need rights? *New Scientist, 234*(3130), 24-25. doi:https://doi.org/10.1016/S0262-4079(17)31163-6
Castanheira, L., Silva, C., Cheniaux, E., & Telles-Correia, D. (2019). Neuroimaging Correlates of Depression—Implications to Clinical Practice. *Frontiers in Psychiatry, 10*(703). doi:10.3389/fpsyt.2019.00703
Chapman, C. R., Mehta, K. S., Parent, B., & Caplan, A. L. (2019). Genetic discrimination: emerging ethical challenges in the context of advancing technology. *Journal of Law and the Biosciences*.
Chaudhary, P., & Agrawal, R. (2018). Emerging threats to security and privacy in brain computer interface. *International Journal of Advanced Studies of Scientific Research, 3*(12).
Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience, 20*(1), 115-125. doi:10.1038/nn.4450
Chizeck, H. J., & Bonaci, T. (2014). Brain-Computer Interface Anonymizer. In: Google Patents.
Clausen, J., Fetz, E., Donoghue, J., Ushiba, J., Spörhase, U., Chandler, J., . . . Soekadar, S. R. (2017). Help, hope, and hype: Ethical dimensions of neuroprosthetics. *Science, 356*(6345), 1338-1339.
Collins, R. F., & Harrington, D. J. (1999). *First Corinthians* (Vol. 7): Liturgical Press.
Committee, U. H. R. (1993). *General Comment No. 22: The Right to Freedom of Thought, Conscience and Religion (Article 18), UN Doc. No*. Retrieved from
Committee on Legal Affairs and Human Rights (2020). The brain-computer interface: new rights or new threats to fundamental freedoms? Report Doc. 15147, 24 September 2020.
Dainton, B., & Bayne, T. (2005). Consciousness as a guide to personal persistence. *Australasian Journal of Philosophy, 83*(4), 549-571.
de Vos, F., Koini, M., Schouten, T. M., Seiler, S., van der Grond, J., Lechner, A., . . . Rombouts, S. A. (2018). A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease. *Neuroimage, 167*, 62-72.
Dennett, D. C. (2015). *Elbow room: The varieties of free will worth wanting*: mit Press.

Douglas, T. (2014). Criminal rehabilitation through medical intervention: moral liability and the right to bodily integrity. *The Journal of Ethics*, *18*(2), 101-122.
Douglas, T., & Forsberg, L. (2021). Three Rationales for a Legal Right to Mental Integrity. *Neurolaw: Advances in Neuroscience, Justice & Security*, 179.

Farah, M. J. (2002). Emerging ethical issues in neuroscience. *Nature neuroscience, 5*(11), 1123-1129.
Farah, M. J., Illes, J., Cook-Deegan, R., Gardner, H., Kandel, E., King, P., . . . Wolpe, P. R. (2004). Neurocognitive enhancement: what can we do and what should we do? *Nature Reviews Neuroscience, 5*(5), 421-425.
Farahany, N. A. (2012). Incriminating thoughts. *Stan. L. Rev., 64*, 351.
Farahany, N. A. (2019). The Costs of Changing Our Minds. *Emory LJ, 69*, 75.
Favaretto, M., De Clercq, E., Gaab, J., & Elger, B. S. (2020). First do no harm: An exploration of researchers' ethics of conduct in Big Data behavioral studies. *Plos one, 15*(11), e0241865.
Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Kübler, A. (2012). P300 brain computer interface: current challenges and emerging trends. *Frontiers in neuroengineering, 5*, 14.
Ferretti, A., & Ienca, M. (2018). Enhanced cognition, enhanced self? On neuroenhancement and subjectivity. *Journal of Cognitive Enhancement, 2*(4), 348-355.
Fins, J. J. (2004). Neuromodulation, free will and determinism: lessons from the psychosurgery debate. *Clinical Neuroscience Research, 4*(1-2), 113-118.
Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass, 2*(2), 347-357.
Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal, 33*(4), 111-117.
Garrett, B. (2002). *Personal identity and self-consciousness*: Routledge.
Gaustad, E. S. (2001). *Roger Williams: Prophet of Liberty*: Oxford University Press.
Gilbert, F., Cook, M., O'Brien, T., & Illes, J. (2019). Embodiment and estrangement: Results from a first-in-human "intelligent BCI" trial. *Science and engineering ethics, 25*(1), 83-96.
Gilbert, F., Goddard, E., Viaña, J. N. M., Carter, A., & Horne, M. (2017). I miss being me: Phenomenological effects of deep brain stimulation. *AJOB neuroscience, 8*(2), 96-109.
Goering, S., Klein, E., Sullivan, L. S., Wexler, A., y Arcas, B. A., Bi, G., . . . Gallant, J. (2021). Recommendations for Responsible Development and Application of Neurotechnologies. *Neuroethics*, 1-22.
Gordon, B. A., Blazey, T. M., Su, Y., Hari-Raj, A., Dincer, A., Flores, S., . . . Xiong, C. (2018). Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. *The Lancet Neurology, 17*(3), 241-250.
Greenberg, A. (2019). Inside the Mind's Eye: An International Perspective on Data Privacy Law in the Age of Brain Machine Interfaces. *Alb. LJ Sci. & Tech., 29*, 79.
Hallinan, D., Schütz, P., Friedewald, M., & de Hert, P. (2014). Neurodata and neuroprivacy: Data protection outdated? *Surveillance & Society, 12*(1), 55-72.

Hayes, B. (2017). Migration and data protection: Doing no harm in an age of mass displacement, mass surveillance and “big data”. *International Review of the Red Cross, 99*(904), 179-209.
Haynes, J.-D. (2011). Brain reading: Decoding mental states from brain activity in humans. *The Oxford Handbook of Neuroethics*, 3-13.
Haynes, J.-D., & Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Current biology, 15*(14), 1301-1307.
Hohfeld, W. N. (1913). Some fundamental legal conceptions as applied in judicial reasoning. *Yale Lj, 23*, 16.
Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural Decoding of Visual Imagery During Sleep. *Science, 340*(6132), 639. doi:10.1126/science.1234330
Hubbard, J., Kikumoto, A., & Mayr, U. (2019). EEG Decoding Reveals the Strength and Temporal Dynamics of Goal-Relevant Representations. *Scientific Reports, 9*(1), 9051. doi:10.1038/s41598-019-45333-6
Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N. Y., Vu, A. T., & Gallant, J. L. (2016). Decoding the Semantic Content of Natural Movies from Human Brain Activity. *Frontiers in Systems Neuroscience, 10*(81). doi:10.3389/fnsys.2016.00081
Ienca, M. (2015). *Neuroprivacy, neurosecurity and brain-hacking: Emerging issues in neural engineering.* Paper presented at the Bioethica Forum.
Ienca, M. (2017). The Right to Cognitive Liberty. *Scientific American, 317*(2), 10-10.
Ienca, M. (2019a). Democratizing cognitive technology: a proactive approach. *Ethics and Information Technology, 21*(4), 267-280.
Ienca, M. (2019b). *Intelligenza2: per un'unione di intelligenza naturale e artificiale*: Rosenberg & Sellier.
Ienca, M., & Andorno, R. (2017a). A new category of human rights: neurorights. Retrieved from http://blogs.biomedcentral.com/bmcblog/2017/04/26/new-category-human-rights-neurorights/
Ienca, M., & Andorno, R. (2017b). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy, 13*(1), 5. doi:10.1186/s40504-017-0050-1
Ienca M., Fins J.J., Jox R.J, Jotterand F., Voeneky S., Andorno R., Ball T., Castelluccia C., Chavarriaga R., Chneiweiss H., Ferretti A., Friedrich O., Hurst S., Merkel G., Molnár-Gábor F., Rickli J.M., Scheibner J., Vayena E., Yuste R., Kellmeyer P. (2021). Towards a Governance Framework for Brain Data. SSRN.
Ienca, M., & Haselager, P. (2016). Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology, 18*(2), 117-129.
Ienca, M., Haselager, P., & Emanuel, E. J. (2018). Brain leaks and consumer neurotechnology. *Nature biotechnology, 36*(9), 805-810.
Ienca, M., & Ignatiadis, K. (2020). Artificial intelligence in clinical neuroscience: methodological and ethical challenges. *AJOB neuroscience, 11*(2), 77-87.
Ienca, M., & Malgieri, G. (2021). Mental Data Protection and the GDPR. Available at SSRN 3840403.
Ienca, M., Shaw, D. M., & Elger, B. (2019). Cognitive enhancement for the ageing world: opportunities and challenges. *Ageing and Society, 39*(10), 2308-2321.
Ienca, M., & Vayena, E. (2018). Cambridge analytica and online manipulation. *Scientific American, 30*.
Illes, J., Kirschen, M. P., & Gabrieli, J. D. (2003). From neuroimaging to neuroethics. *Nature neuroscience, 6*(3), 205-205.

Illes, J., Rosen, A. C., Huang, L., Goldstein, R., Raffin, T. A., Swan, G., & Atlas, S. (2004). Ethical consideration of incidental findings on adult brain MRI in research. *Neurology, 62*(6), 888-890.
Jobsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science, 198*(4323), 1264-1267.
Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*(7185), 352-355.
Kober, H., Lacadie, C. M., Wexler, B. E., Malison, R. T., Sinha, R., & Potenza, M. N. (2016). Brain activity during cocaine craving and gambling urges: an fMRI study. *Neuropsychopharmacology, 41*(2), 628-637.
Koh, H. H. (1998). How is international human rights law enforced. *Ind. lJ, 74*, 1397.
Lauterbur, P. C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature, 242*(5394), 190-191.
Lavazza, A. (2018). Freedom of Thought and Mental Integrity: The Moral Requirements for Any Neural Prosthesis. *Frontiers in Neuroscience, 12*(82). doi:10.3389/fnins.2018.00082
Lefaucheur, J.-P., André-Obadia, N., Antal, A., Ayache, S. S., Baeken, C., Benninger, D. H., . . . De Ridder, D. (2014). Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rTMS). *Clinical Neurophysiology, 125*(11), 2150-2206.
Levy, N. (2007). Neuroethics: Challenges for the 21st century. Cambridge University Press.
Li, M., Xu, H., & Lu, S. (2018). Neural basis of depression related to a dominant right hemisphere: A resting-state fMRI study. *Behavioural neurology, 2018*.
Long, A. A. (2002). *Epictetus: A Stoic and Socratic guide to life*: Clarendon Press.
Luzzatti, L. (2006). *God in Freedom: Studies in the Relations Between Church and State*: Cosimo, Inc.
Marshall, J. (2014). *Human rights law and personal identity*: Routledge.
Marshel, J. H., Kim, Y. S., Machado, T. A., Quirin, S., Benson, B., Kadmon, J., . . . Deisseroth, K. (2019). Cortical layer-specific critical dynamics triggering perception. *Science, 365*(6453). doi:10.1126/science.aaw5202
McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44*(2), 379-387.
Meslin, E. M. (1990). Protecting human subjects from harm through improved risk judgments. *IRB: Ethics & Human Research, 12*(1), 7-10.
Mill, J. S. (1859). On liberty In: London: Thinker's library.
Milton, J. (1791). *Comus: A Mask*: proprietors, under the direction of John Bell.
Moreno, J. D. (2003). Neuroethics: an agenda for neuroscience and society. *Nature Reviews Neuroscience, 4*(2), 149-153.
Moscrop, H. (2014). Enforcing International Human Rights Law: Problems and Prospects. In.
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:10.1038/s41467-019-10994-4
Munoz, J. M. (2019). Chile--right to free will needs definition. *Nature, 574*(7780), 634-635.

Nawrot, O. (2019). What about the Interior Castle? Response to Ienca's and Andorno's New Human Rights in the Age of Neuroscience and Neurotechnology. *Roczniki Teologiczne, 66*(3), 69-85.

Nickel, J., Pogge, T., Smith, M., & Wenar, L. (2013). Stanford Encyclopedia of Philosophy, Human Rights. In: December.

Noonan, H. W. (2019). *Personal identity*: Routledge.

Nussbaum, M. (2007). Human rights and human capabilities. Harv. Hum. Rts. J., 20, 21.

Nuyujukian, P., Albites Sanabria, J., Saab, J., Pandarinath, C., Jarosiewicz, B., Blabe, C. H., . . . Simeral, J. D. (2018). Cortical control of a tablet computer by people with paralysis. *Plos one, 13*(11), e0204566.

OECD-Council. (2019). *OECD Recommendation on Responsible Innovation in Neurotechnology*. Retrieved from https://www.oecd.org/science/recommendation-on-responsible-innovation-in-neurotechnology.htm

Omurtag, A., Aghajani, H., & Keles, H. O. (2017). Decoding human mental states by whole-head EEG+fNIRS during category fluency task performance. *J Neural Eng, 14*(6), 066003. doi:10.1088/1741-2552/aa814b

Palaniappan, R., & Mandic, D. P. (2007). Biometrics from brain electrical activity: A machine learning approach. *IEEE transactions on pattern analysis and machine intelligence, 29*(4), 738-742.

Palk, A., Illes, J., Thompson, P. M., & Stein, D. J. (2020). Ethical issues in global neuroimaging genetics collaborations. *NeuroImage*, *221*, 1172

Pereboom, D., & Caruso, G. (2002). Hard-incompatibilist existentialism: Neuroscience, punishment, and meaning in life. *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience, OUP Editors, Gregg D. Caruso and Owen Flanagan, Forthcoming*.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron, 72*(5), 692-697. doi:10.1016/j.neuron.2011.11.001

Polenberg, R. (1996). Cardozo and the Criminal Law: Palko v. Connecticut Reconsidered. *J. Sup. Ct. Hist.*, 92.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310*(5756), 1963-1966.

Powell, C., Munetomo, M., Schlueter, M., & Mizukoshi, M. (2013). *Towards thought control of next-generation wearable computing devices.* Paper presented at the International conference on brain and health informatics.

Price-Williams, D. (1957). Proprioception and personal identity. *Philosophy and Phenomenological Research, 17*(4), 536-545.

Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*: OUP Oxford.

Pugh, J., Pycroft, L., Sandberg, A., Aziz, T., & Savulescu, J. (2018). Brainjacking in deep brain stimulation and autonomy. *Ethics and Information Technology, 20*(3), 219-232.

Pycroft, L., Boccard, S. G., Owen, S. L., Stein, J. F., Fitzgerald, J. J., Green, A. L., & Aziz, T. Z. (2016). Brainjacking: implant security issues in invasive neuromodulation. *World neurosurgery, 92*, 454-462.

Rainey, S., Bublitz, J. C., Maslen, H., & Thornton, H. (2019). Data as a cross-cutting dimension of ethical importance in direct-to-consumer neurotechnologies. *AJOB neuroscience, 10*(4), 180-182.
Rainey, S., McGillivray, K., Akintoye, S., Fothergill, T., Bublitz, C., & Stahl, B. (2020). Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology? *J Law Biosci*. doi:10.1093/jlb/lsaa051
Reider, L. (1998). Toward a new test for the insanity defense: Incorporating the discoveries of neuroscience into moral and legal theories. *UCLA L. Rev., 46*, 289.
Reinhart, R. M., & Nguyen, J. A. (2019). Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nature neuroscience, 22*(5), 820-827.
Repetti, R. (2018). *Buddhism, Meditation, and Free Will: A Theory of Mental Freedom*: Routledge.
Rickli, J.-M., & Ienca, M. (2021). The Security and Military Implications of Neurotechnology and Artificial Intelligence. *Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications*, 197-214.
Roskies A. Neuroethics for the new millennium. Neuron. 2002;35:21–3.
Safire, W. (2002). *Visions for a new field of" neuroethics".* Paper presented at the Neuroethics: Mapping the Field, Conference Proceedings, May 13-14, 2002, San Francisco, California.
Safron, A., Klimaj, V., Sylva, D., Rosenthal, A., Li, M., Walter, M., & Bailey, J. M. (2018). Neural correlates of sexual orientation in heterosexual, bisexual, and homosexual women. *Scientific Reports, 8*(1), 1-14.
Schaffer, J. (2015). What not to multiply without necessity. *Australasian Journal of Philosophy, 93*(4), 644-664.
Schechtman, M. (1994). The same and the same: Two views of psychological continuity. *American Philosophical Quarterly, 31*(3), 199-212.
Scott, G. (2013). The neurotechnology revolution has arrived. The Futurist, 47(5), 6.
Schreiber, D., Fonzo, G., Simmons, A. N., Dawes, C. T., Flagan, T., Fowler, J. H., & Paulus, M. P. (2013). Red brain, blue brain: Evaluative processes differ in Democrats and Republicans. *Plos one, 8*(2), e52970.
Schwarz, C. G., Kremers, W. K., Therneau, T. M., Sharp, R. R., Gunter, J. L., Vemuri, P., . . . Jack, C. R. (2019). Identification of Anonymous MRI Research Participants with Face-Recognition Software. *New England Journal of Medicine, 381*(17), 1684-1686. doi:10.1056/NEJMc1908881
Scott, G. (2013). The neurotechnology revolution has arrived. *The Futurist*, *47*(5), 6.
Sententia, W. (2004). Neuroethical considerations: cognitive liberty and converging technologies for improving human cognition. *Annals of the New York Academy of Sciences, 1013*(1), 221-228.
Shen, F. X. (2013). Neuroscience, mental privacy, and the law. *Harv. JL & Pub. Pol'y, 36*, 653.
Shen, F. X. (2016). The overlooked history of neurolaw. *Fordham L. Rev., 85*, 667.
Smith, C. M. (2005). Origin and Uses of Primum Non Nocere—Above All, Do No Harm! *The Journal of Clinical Pharmacology, 45*(4), 371-377. doi:https://doi.org/10.1177/0091270004273680
Soon, C. S., He, A. H., Bode, S., & Haynes, J.-D. (2013). Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences, 110*(15), 6217-6222.

Spence, S. (1996). Free will in the light of neuropsychiatry. *Philosophy, Psychiatry, & Psychology, 3*(2), 75-90.
Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, *8*(2).
Taylor, J. S., Harp, J. A., & Elliott, T. (1991). Neuropsychologists and neurolawyers. *Neuropsychology, 5*(4), 293.
Thompson, K. (2019). Committing Crimes with BCIs: How Brain-Computer Interface Users can Satisfy Actus Reus and be Criminally Responsible. *Neuroethics*, 1-12.
Thomson, J. J. (1990). *The realm of rights*: Harvard University Press.
Turner, D. C., & Sahakian, B. J. (2006). Neuroethics of cognitive enhancement. *BioSocieties, 1*(1), 113-123.
Van Inwagen, P. (1997). Materialism and the psychological-continuity account of personal identity. *Philosophical Perspectives, 11*, 305-319.
Vidal, J. J., & JJ, V. (1973). Toward direct brain-computer communication.
Welford, A. T. (1970). Mental Integrity and the Nature of Life. *Medical Journal of Australia, 1*(23), 1135-1138. doi:https://doi.org/10.5694/j.1326-5377.1970.tb84477.x
Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review, 25*(1), 166.
Wexler, A., & Reiner, P. B. (2019). Oversight of direct-to-consumer neurotechnologies. *Science, 363*(6424), 234. doi:10.1126/science.aav0223
Wolpe, P. R. (2017). Neuroprivacy and Cognitive Liberty. In *The Routledge Handbook of Neuroethics* (pp. 214-224): Routledge, Taylor & Francis Group.
Woolf, V. (1929). *A room of one's own and three guineas*: OUP Oxford.
Yuste, R., Genser, J., & Herrmann, S. (2021). It’s Time for Neuro-Rights. *Horizons: Journal of International Relations and Sustainable Development*(18), 154-165.
Yuste, R., Goering, S., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., . . . Illes, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature News, 551*(7679), 159.
Zeki, S., Goodenough, O., & O'Hara, E. A. (2004). How neuroscience might advance the law. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1451), 1677-1684.