

‘Neurorights’

*A Human Rights–Based Approach for Governing Neurotechnologies**Philipp Kellmeyer*

I. INTRODUCTION

The combination of digital technologies for data collection and processing with advances in neurotechnology promises a new generation of highly adaptable, AI-based brain–computer interfaces for clinical but also consumer-oriented purposes. By integrating various types of personal data – physiological data, behavioural data, biographical data, and other types – such systems could become adept at inferring mental states and predicting behaviour, for example, for intended movements or consumer choices. This development has spawned a discussion – often framed around the idea of ‘neurorights’ – around how to protect mental privacy and mental integrity in the interaction with AI-based systems. Here, I review the current state of this debate from the perspective of philosophy, ethics, neuroscience, and psychology and propose some conceptual refinements on how to understand mental privacy and mental integrity in human–AI interactions.

The dynamic convergence of neuroscience, neurotechnology, and AI that we see today was initiated by progress in the scientific understanding of brain processes, the invention of computing machines and algorithmic programming in the early and mid-twentieth century.

In his book *The Sciences of the Artificial*, computer science, cybernetics, and AI pioneer *Herbert A. Simon* characterizes the relationship between the human mind and the human brain as follows:

As our knowledge increases, the relation between physiological and information-processing explanations will become just like the relation between quantum-mechanical and physiological explanations in biology (or the relation between solid-state physics and programming explanations in computer science). They constitute two linked levels of explanation with (in the case before us) the limiting properties of the inner system showing up at the interface between them.¹

This description captures the general spirit and prevailing analogy of the beginnings and early decades of the computer age: just as the computer is the hardware on which software is implemented, the brain is the hardware on which the mind runs. In the early 1940s, well before the first digital computers were built, *Warren S. McCulloch* and *Walter Pitts* introduced the idea of artificial neural networks that could compute logical functions.² Later, in 1950, *Donald Hebb*

¹ HA Simon, *The Sciences of the Artificial* (2001) 83.

² WS McCulloch and W Pitts, ‘A Logical Calculus of the Ideas Immanent in Nervous Activity’ (1943) 5(4) *The Bulletin of Mathematical Biophysics* 115–133 <https://doi.org/10.1007/BF02478259>.

in *The Organization of Behavior*³ developed a theory of efficient encoding of statistics in neural networks which became a foundational text for early AI researchers and engineers. Later yet, in 1958, *Frank Rosenblatt* introduced the concept of a perceptron, a simple artificial neural network, which had comparatively limited information-processing capabilities back then but constitutes the conceptual basis from which the powerful artificial neural networks for deep learning are built today.

Much of this early cross-fertilization between discoveries in neurophysiology and the design of computational systems was driven by the insight that both computers and human brains can be broadly characterized as information-processing systems. This analogy certainly has intuitive appeal and motivates research programs to this day. The aim is to find a common framework that unifies approaches from diverse fields – computer science, AI, cybernetics, cognitive science, neuroscience – into a coherent account of information processing in (neuro)biological and artificial systems. But philosophy, especially philosophy of mind, (still) has unfinished business and keeps throwing conceptual wrenches – in the form of thought experiments, the most famous of which is arguably *John Searle’s* Chinese Room Argument⁴ – into this supposedly well-oiled machine of informational dualism.

Today, through the ‘super-convergence’⁵ of digital and information technologies, this original affinity and mutual inspiration between computer science (artificial neural networks, cognitive systems, and other approaches) and the sciences of the human brain and cognition is driving a new generation of AI-inspired neurotechnology and neuroscience-inspired AI.⁶

In the field of brain–computer interfacing, for example, the application of AI-related machine learning methods, particularly artificial neural networks for deep learning, have demonstrated superior performance to conventional algorithms.⁷ The same machine learning approach also excels in distinguishing normal from disease-related patterns of brain activity, for example, in finding patterns of epileptic brain activity in conventional electroencephalography (EEG) diagnostics.⁸ These and other successes in applying AI-related methods to analysing and interpreting brain data drives an innovation ecosystem in which not only academic researchers and private companies, but also military research organizations invest heavily (and compete) in the field of ‘intelligent’ neurotechnologies.⁹ This development has spawned an increasing number of analyses and debates on the ethical, legal, social, and policy-related relevance of

³ DO Hebb, *The Organization of Behavior* (1949).

⁴ JR Searle, ‘Minds, Brains, and Programs’ (1980) 3 *Behavioral and Brain Sciences* 417–457 <https://doi.org/10.1017/S0140525X00005756>.

⁵ The confluence of big data, artificial neural networks for deep learning, the web, microsensors, and other transformative technologies, cf. H Hahn and A Schreiber, ‘E-Health’ in R Neugebauer (ed), *Digital Transformation* (2019) 311–334 https://doi.org/10.1007/978-3-662-58134-6_19.

⁶ P Kellmeyer, ‘Artificial Intelligence in Basic and Clinical Neuroscience: Opportunities and Ethical Challenges’ (2019) 25(4) *Neuroforum* 241–250 <https://doi.org/10.1515/nf-2019-0018>; AH Marblestone, G Wayne, and KP Kording, ‘Toward an Integration of Deep Learning and Neuroscience’ (*Frontiers in Computational Neuroscience*, 14 September 2016) 94 <https://doi.org/10.3389/fncom.2016.00094>.

⁷ D Kuhner and others, ‘A Service Assistant Combining Autonomous Robotics, Flexible Goal Formulation, and Deep-Learning-Based Brain–Computer Interfacing’ (2019) 116 *Robotics and Autonomous Systems* 98–113 <https://doi.org/10.1016/j.robot.2019.02.015>; F Burget and others, ‘Acting Thoughts: Towards a Mobile Robotic Service Assistant for Users with Limited Communication Skills’ (*IEEE*, 9 November 2017) 1–6 <https://doi.org/10.1109/ECMR.2017.8098658>.

⁸ LAW Gemein and others, ‘Machine-Learning-Based Diagnostics of EEG Pathology’ (2020) 220 *NeuroImage* 117021 <https://doi.org/10.1016/j.neuroimage.2020.117021>.

⁹ P Kellmeyer, ‘Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices’ (2018) 14 *Neuroethics* 83–98 <https://doi.org/10.1007/s12152-018-9371-x> (hereafter Kellmeyer, ‘Big Brain Data’); M Ienca, P Haselager, and EJ Emanuel, ‘Brain Leaks and Consumer Neurotechnology’ (2018) 36 *Nature Biotechnology* 805–810 <https://doi.org/10.1038/nbt.4240>.

brain data analytics and intelligent neurotechnologies.¹⁰ Central concepts in this debate are the notions of mental privacy and mental integrity.

In this chapter, I will first give an account of the current understanding as well as ethical and legal implications of mental privacy and propose some conceptual refinements. Then I will attempt to clarify the conceptual foundations of mental integrity and propose a description that can be applied across various contexts. I will then address the debate on neurorights and advocate for an intermediate position between human rights conservatism (no new rights are necessary to protect mental privacy and integrity) and human rights reformism (existing human rights frameworks are insufficient to protect mental privacy and integrity and need to be revised). I will argue that the major problem is not the lack of well-conceptualized fundamental rights but insufficient pathways and mechanisms for applying these rights to effectively protect mental privacy and mental integrity from undue interference.

II. MENTAL PRIVACY

1. *The Mental Realm: The Spectre of Dualism, Freedom of Thought and Related Issues*

As outlined in the introduction and in the absence of a universal definition, I propose the following pragmatic operational description: ‘Mental privacy denotes the domain of a person’s active brain processes and experiences – perceptions, thoughts, emotions, volition; roughly corresponding to *Kant’s* notion of the *locus internus* in philosophy¹¹ – which are exceptionally hard (if not impossible) to access externally.’ The mental ‘realm’ implicated in this description refers to an agent’s phenomenological subjective experiences, indicated in language by terms such as ‘thoughts’, ‘inner speech’, ‘intentions’, ‘beliefs’, and ‘desires’, but also ‘fear’, ‘anxiety’ and emotions (such as ‘sadness’). While it makes intuitive sense, from a folk-psychological perspective, calling for special protection to this mental realm is predicated on a precise understanding of the relationship between levels of subjective experiences and corresponding brain processes – a requirement that neuroscientific evidence and models cannot meet¹².

From a monist and materialist position, these qualitative terms offer convenient ways for us to refer to subjective experiences, insisting that there is – in the strict ontological sense – nothing but physical processes in the human body (and the brain most of all), no dualistic ‘second

¹⁰ P Kellmeyer and others, ‘Neuroethics at 15: The Current and Future Environment for Neuroethics’ (2019) 10(3) *AJOB Neuroscience* 104–110; S Rainey and others, ‘Data as a Cross-Cutting Dimension of Ethical Importance in Direct-to-Consumer Neurotechnologies’ (2019) 10(4) *AJOB Neuroscience* 180–182 <https://doi.org/10.1080/21507740.2019.1665134>; Kellmeyer, ‘Big Brain Data’ (n 9); R Yuste and others, ‘Four Ethical Priorities for Neurotechnologies and AI’ (2017) 551 (7679) *Nature News* 159 <https://doi.org/10.1038/551159a> (hereafter Yuste and others, ‘Four Ethical Priorities for Neurotechnologies and AI’); M Ienca and R Andorno, ‘Towards New Human Rights in the Age of Neuroscience and Neurotechnology’ (2017) 13 *Life Sciences, Society and Policy* 5 <https://doi.org/10.1186/s40504-017-0050-1> (hereafter Ienca and Andorno, ‘Towards New Human Rights in the Age of Neuroscience and Neurotechnology’).

¹¹ I Leclerc, ‘The Meaning of “Space” in LW Beck (ed), *Kant’s Theory of Knowledge: Selected Papers from the Third International Kant Congress* (1974) 87–94 https://doi.org/10.1007/978-94-010-2294-1_10. This division into a *locus internus* (as described here) and *locus externus* – the set of externally observable facts about human behavior – is reflected in the ongoing debate about the nature of human phenomenological experience, consciousness, and free will in philosophy; the intricacies and ramifications of which lie outside of the scope of this article. For recent contributions to these overlapping debates, see e.g. the excellent overview in P Goff’s *Galileo’s Error* (2020).

¹² I deliberately refrain from qualifying this statement as to whether, and if so when, we should expect neuroscience to ever be able to give a full account of a mechanistic understanding, both for conceptual reasons and practical reasons, for example, inherent limitations of current, and likely future, measurement tools in observing brain processes at the ‘right’ levels of granularity or scale (microscale, mesoscale, and macroscale) and at the appropriate level of temporal and frequency-related sampling to relate them to any given subjective experience.

substance’ or, as *René Descartes* referred to it, *mens rea*. In such an interpretation, there is no ‘mind-body problem’ because there is no such thing as a mind to begin with and the human practice of talking as if there was a mental realm that is separate from the physical realm arises from our (again folk-psychological, or anthropological) propensity to interpret our subjective experience as separate from brain processes, perhaps because we have no direct sensory access to these processes in the first place.

This spectre of dualism, the illusion – as a materialist (e.g. a physicalist) would put it – that our physical brain processes and our experiences are separate ‘things’, is so convincing and persuasive that it not only haunts everyday language, but is also deeply engrained in concept-formation and theorizing in psychological and neuroscientific disciplines such as experimental psychology or cognitive neuroscience as well as the medical fields of neurology, psychosomatic medicine, and psychiatry.¹³

To date, there is no widely accepted and satisfying explanation of the precise relationship between the phenomenological level of subjective experience and brain processes. This conundrum allows for a wide range of theoretical positions, from strictly neuroessentialist and neurodeterministic interpretations (i.e. there is nothing separate from brain processes; and brain activity does not give rise to but simply *is* nothing but neurophysiology), to positions that emphasize the ‘4E’¹⁴ character of human cognition and all the way to modern versions of dualist positions, such as ‘naturalistic dualism’¹⁵. An interesting intermediate position that has experienced somewhat of a renaissance in the philosophy of mind in recent years is the concept of panpsychism. The main idea in panpsychism is that consciousness is a fundamental and ubiquitous feature of the natural world. In this view, the richness of our mental experience could be explained as an emerging property that depends on the complexity of biological organisms and their central nervous systems.¹⁶ Intriguingly, there seem to be conceptually rich connections between advanced neuroscientific theories of consciousness, particularly the so-called Integrated Information Theory (IIT)¹⁷, and emergentist panpsychist interpretations of consciousness and mental phenomena.¹⁸ The reason why this is relevant for our topic here – brain data, information about brain processes, and neurotechnology – is that these conceptual and neuroscientific advances in building a unified theory of causal mechanisms of subjective experience might become an important tenet for future analytical approaches to decoding brain data from neurotechnologies and inferring mental information from these analyses.

¹³ Consider, for example, the concept of ‘dissociation’ in psychiatry (in the context of post-traumatic stress disorder) or neurology (in epilepsy), the notion that brain processes and mental processes can become uncoupled.

¹⁴ The 4E framework emphasizes that human cognition cannot be separated from the way in which cognitive processes are embodied (in a physical body [German: ‘*Leib*’]), embedded (into the environment), extended (how we use tools to facilitate cognition), and enactive (cognition enacts itself in interaction with others) R Menary, ‘Introduction to the Special Issue on 4E Cognition’ (2010) 9(4) *Phenomenology and the Cognitive Sciences* 459–463 <https://doi.org/10.1007/s11097-010-9187-6>.

¹⁵ D Chalmers, ‘Naturalistic Dualism’ in S Schneider and M Velmans (eds), *The Blackwell Companion to Consciousness* (2017) 363–373 <https://doi.org/10.1002/9781119132363.ch26>.

¹⁶ P Goff, *Consciousness and Fundamental Reality* (2017); P Goff, W Seager, and S Allen-Hermanson, ‘Panpsychism’ in EN Zalta (ed), *The Stanford Encyclopedia of Philosophy* (2020) <https://plato.stanford.edu/archives/sum2020/entries/panpsychism/>.

¹⁷ G Tononi and others, ‘Integrated Information Theory: From Consciousness to Its Physical Substrate’ (2016) 17(7) *Nature Reviews Neuroscience* 450–461 <https://doi.org/10.1038/nrn.2016.44>.

¹⁸ HH Mørch, ‘Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism?’ (2019) 84(5) *Erkenntnis* 1065–1085 <https://doi.org/10.1007/s10670-018-9995-6>.

2. Privacy of Data and Information: Ownership, Authorship, Interest, and Responsibility

Before delving into the current debate around mental privacy, let me provide a few propaedeutical thoughts on the terminology and conceptual foundations of privacy and how it is understood in the context of data and information processing. Etymologically, ‘privacy’ originates from the Latin term *privatus* which means ‘withdrawn from public life’.¹⁹ An important historical usage context for the concept of ‘privacy’ was in the military and warfare domain, for example in the notion of ‘privateers’, that is, a person or ship that privately participated in an armed naval conflict under official commission of war (distinguishing privateering from outlawed activities such as piracy).²⁰ The term and concept has a rich history in jurisprudence and the law. Lacking the space to retrace all ramifications of the legal-philosophical understandings of privacy, one notion that seems relevant for our context here – and that sets privacy apart from the related notion of seclusion and secrecy²¹ – is that privacy ultimately concerns a person’s ‘autonomy within society’.²² In the current age of digital information technology, this autonomy extends into the realm of the informational—in other words, the ‘infosphere’ as elucidated by Luciano Floridi²³—which is reflected by an increasing number of ethical and legal analyses of ‘informational privacy’ and the metamorphosis of persons into ‘data subjects’ and digital service providers into ‘data controllers’ in the digital realm.²⁴ In this context, it may be worthwhile to remind us that data and information (and knowledge for that matter), though intricately intertwined, are not interchangeable notions. Whereas data are ‘numbers and words without relationships’, information are ‘numbers and words with relationships’ and knowledge refers to inferences gleaned from information.²⁵ This distinction is important for the development and application of granular and context-sensitive legal and policy instruments for protecting a person’s privacy.²⁶

For contexts in which questions around the protection of (and threats to) data or informational privacy are originating from the creation, movement, storage and analysis of digital data, it would seem appropriate to conceptualize ‘informational privacy’ as: autonomy of persons over the collection, access and use of data and information about themselves. Related to these questions, this expanding discussion has made the question of data (and information) ownership

¹⁹ TF Hoad, ‘Private’ in TF Hoad (ed), *The Concise Oxford Dictionary of English Etymology* (2003) www.oxfordreference.com/view/10.1093/acref/9780192830982.001.0001/acref-9780192830982-e-11928.

²⁰ Another legacy in the military domain is the rank of private, i.e. soldiers of the lowest military rank.

²¹ See e.g. the usage definition from Merriam Webster, ‘Privacy’ (*Merriam Webster Dictionary*) www.merriam-webster.com/dictionary/privacy.

²² J Hirshleifer, ‘Privacy: Its Origin, Function, and Future’ (1980) 9(4) *The Journal of Legal Studies* 649–664.

²³ L Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality* (2014).

²⁴ AD Vanberg, ‘Informational Privacy Post GDPR: End of the Road or the Start of a Long Journey?’ (2021) 25(1) *The International Journal of Human Rights* 52–78 <https://doi.org/10.1080/13642987.2020.1789109> (hereafter Vanberg, ‘Informational Privacy Post GDPR’); TW Kim and BR Routledge, ‘Informational Privacy, A Right to Explanation, and Interpretable AI’ in IEEE (ed), 2018 *IEEE Symposium on Privacy-Aware Computing (PAC)* (2018) 64–74 <https://doi.org/10.1109/PAC.2018.00013>; AD Moore, ‘Toward Informational Privacy Rights 2007 Editor’s Symposium’ (2007) 44(4) *San Diego Law Review* 809–846; L Floridi, ‘Four Challenges for a Theory of Informational Privacy’ (2006) 8(3) *Ethics and Information Technology* 109–119 <https://doi.org/10.1007/s10676-006-9121-3> (hereafter Floridi, ‘Four Challenges for a Theory of Informational Privacy’).

²⁵ J Pohl, ‘Transition From Data to Information’ in *Collaborative Agent Design Research Center Technical Report - RESU72* (2001) 1–8.

²⁶ Depending on the context, a very different granularity of privacy protection might be necessary. Consider, for example, the difference between collecting only one specific type of biometric data (without other contextual data) vs. collecting multimodal personal data to glean health-related information in a consumer technology context, which would require different granularity of data and information protection.

a central aspect of ethical and legal scholarship and policy debates.²⁷ In a legal context, the protection of data or informational privacy are relevant, *inter alia*, in trade law (e.g. confidential trade secrets), copyright law, health law and many other legal areas. Importantly, however, individuals do not have property rights regarding their personal information, e.g. information about their body, health and disease in medical records.²⁸ Separate from the question of ownership of personal information is the question of authorship, in other words, who can be regarded as the creator of specific data and information about a person.²⁹ But, even in contexts in which persons are neither the author/creator nor the owner of data and information about themselves, they nevertheless have legitimate interests in protecting this information from being misused to their disadvantage, and therefore legitimate interest, and derived thereof, right, to keep it private. This right to informational privacy is now a fundamental tenet in consumer protection laws as well as data protection and privacy laws, for example the European Union’s (EU) General Data Protection Regulation (GDPR).³⁰

Finally, these questions of ownership, authorship, and interests in personal data and information – and the legal mechanisms for protecting the right to informational privacy – of course also raise the questions of responsibility for and stewardship of personal data and information to protect them from unwarranted access and from misuse. Typically, many different participants and stakeholders are involved in the creation, administration, distribution, and use of personal data and information (i.e. the creator(s)/author(s), owner(s), persons with legitimate and vested interests). Under many circumstances, this creates a problem of ascribing responsibility for data stewardship – a diffusion of responsibility. This may be further complicated by the fact that the creator of a particular set of personal information, the owner (and the person to whom these data and information pertain), may reside in different jurisdictions and may therefore be accountable to different data protection and privacy laws.

3. *Mental Privacy: Protecting Data and Information about the Human Brain and Associated Mental Phenomena*

In the debate around ‘neurorights’ the term mental privacy has established itself to refer to the ‘mental realm’ outlined above. However, from a materialist, neurodeterministic position, it would not make much sense to give mental phenomena special juridical protection if we neither have ways to measure these phenomena nor a model of causal mechanisms to give an account of how they arise. For the law, however, such a strict mechanistic interpretation of

²⁷ A Ballantyne, ‘How Should We Think about Clinical Data Ownership?’ (2020) 46(5) *Journal of Medical Ethics* 289–294 <https://doi.org/10.1136/medethics-2018-105340>; P Hummel, M Braun and P Dabrock, ‘Own Data? Ethical Reflections on Data Ownership’ (2020) *Philosophy & Technology* 1–28 <https://doi.org/10.1007/s13347-020-00404-9>; M Mirchev, I Mircheva and A Kerekovska, ‘The Academic Viewpoint on Patient Data Ownership in the Context of Big Data: Scoping Review’ (2020) 22(8) *Journal of Medical Internet Research* <https://doi.org/10.2196/22214>; N Duch-Brown, B Martens and F Mueller-Langer, ‘The Economics of Ownership, Access and Trade in Digital Data’ (SSRN, 17 February 2017), <https://doi.org/10.2139/ssrn.2914144>.

²⁸ Canada Supreme Court, *McInerney v MacDonald* (11 June 1992) 93 Dominion Law Reports 415–31.

²⁹ JC Wallis and CL Borgman, ‘Who Is Responsible for Data? An Exploratory Study of Data Authorship, Ownership, and Responsibility’ (2011) 48(1) *Proceedings of the American Society for Information Science and Technology* 1–10 <https://doi.org/10.1002/meet.2011.14504801188>.

³⁰ Vanberg, ‘Informational Privacy Post GDPR’ (n 24); FT Beke, F Eggers, and PC Verhoef, ‘Consumer Informational Privacy: Current Knowledge and Research Directions’ (2018) 11(1) *Foundations and Trends(R) in Marketing* 1–71; HT Tavani, ‘Informational Privacy: Concepts, Theories, and Controversies’ in KH Himma and HT Tavani (eds), *The Handbook of Information and Computer Ethics* (2008) 131–64 <https://doi.org/10.1002/9780470281819.ch6>; Floridi, ‘Four Challenges for a Theory of Informational Privacy’ (n 24).

mental mechanisms might not be required to ensure adequate protections. Consider, for example, that crimes with large immaterial components such as ‘hate speech’ or ‘perjury’ also contain a large component of internal processes that might remain hidden from the eye of the law. In hate speech, for instance, both the level of internal motivation of the perpetrator as well as the level of internal processes of psychological injury in the injured party do not need to be objectivated in order to establish whether or not a punishable crime was committed.

The precise understanding and interpretation of mental privacy also differs substantially across literatures, contexts, and debates. In legal philosophy, for instance, mental privacy is mainly discussed in the context of foundational questions and justifications in criminal justice such as the concept of *mens rea* (the ‘guilty mind’),³¹ freedom of the will, the feasibility of lie detection, and other ‘neurolaw’ issues.³² In neuroethics, mental privacy is often invoked in discussions around brain data governance and regulation as well as in reference to ‘neurorights’: the question of whether the protection of mental privacy is (or shall become) a part of human rights frameworks and legislation.³³ The discussion here shall be concerned with the latter context.

III. MENTAL INTEGRITY THROUGH THE LENS OF VULNERABILITY ETHICS

Mental integrity, much like the term mental privacy, has an evocative appeal which allows for an intuitive and immediate approximate understanding: to protect the intactness and inviolacy of brain structure and functions (and the associated mental experiences).

Like mental privacy, however, mental integrity is currently still lacking a broadly accepted definition across philosophy, ethics, cognitive science, and neuroscience.³⁴ Most operational descriptions refer to the idea that the structure and function of the human brain and the corresponding mental experiences allow for an integrated mental experience for an individual and that external interference with this integrated experience requires a reasonable justification (such as medication for disturbed states of mind in psychosis, for example) to be morally (and legally) acceptable. The problem that the nature of subjective mental experience, phenomenal consciousness, is inaccessible both internally (as the subject can only describe the qualitative aspects of the experience itself, but not the mechanics of its composite nature) and externally, also affects the way in which we conceptualize the notion of an integrated mind. As an individual – the indivisible person in the literal sense – we mostly experience the world in a more or less unified way, even though separate parallel perceptual, cognitive, and emotive processes have to be integrated in a complex manner to allow for this holistic experience. When being asked, for example, by a curious experimental psychologist or cognitive scientist, to describe the nature of our experience, for example seeing a red apple on a table, we can identify qualitative characteristics of the apple: its shape, texture, colour, and perhaps smell. Yet, we have no shared terminology to describe the quality of our inner experience of seeing the apple – outside of associating particular thoughts, memories, or emotions with this instance of an apple or apples in general.

³¹ P Kellmeyer, ‘Ethical and Legal Implications of the Methodological Crisis in Neuroimaging’ (2017) 26(4) *Cambridge Quarterly of Healthcare Ethics*: CQ: *The International Journal of Healthcare Ethics Committees* 530–554 <https://doi.org/10.1017/S096318011700007X>.

³² G Meynen, ‘Neurolaw: Neuroscience, Ethics, and Law. Review Essay’ (2014) 17(4) *Ethical Theory and Moral Practice* 819–829 <http://www.jstor.org/stable/24478606>; TM Spranger, ‘Neurosciences and the Law: An Introduction’ in TM Spranger (ed), *International Neurolaw* (2012) 1–10 https://doi.org/10.1007/978-3-642-21541-4_1.

³³ Yuste and others, ‘Four Ethical Priorities for Neurotechnologies and AI’ (n 10); Ienca and Andorno, ‘Towards New Human Rights in the Age of Neuroscience and Neurotechnology’ (n 10); Kellmeyer, ‘Big Brain Data’ (n 9).

³⁴ A Lavazza, ‘Freedom of Thought and Mental Integrity: The Moral Requirements for Any Neural Prosthesis’ (*Froniers in Neuroscience*, 19 February 2018) 12 <https://doi.org/10.3389/fnins.2018.00082>.

Put in another way: We all know intuitively what a unified or integrated experience of seeing an apple is like but we cannot explain it in such a way that the descriptions necessarily evoke the same experience(s) in others. To better understand what an integrated experience is like, we might also consider what a disintegrated, disunified, or fragmented experience is like. In certain dream-like states, pathogenic states like psychosis or under the influence of psychoactive substances, an experience can disintegrate into certain constitutive components (e.g. perceiving the shape and colour of the apple separately, yet, simultaneously) or perceptions can be qualitatively altered in countless ways (consider, for instance, the phenomenon of synaesthesia, ‘seeing’ tones or ‘hearing’ colours). This demonstrated potential for the composite nature of mental experiences suggests that it is not inconceivable that we might find more targeted and precise ways to influence the qualitative nature (and perhaps content) of our mental experiences, for example, through precision drugs or neurotechnological interventions.³⁵ Emerging techniques such as optogenetics, for instance, have already been demonstrated to be able to ‘incept’ false memories into a research animal’s brain.³⁶ But our mental integrity can also be compromised by non-neurotechnological interventions of course. Consider approaches from (behavioral) psychology such as nudging or subliminal priming (and related techniques)³⁷ that can influence decision making and choice (and have downstream effects on the experiences associated with these decisions and choices) or more overt psychological interventions such as psychotherapy or the broad – and lately much questioned (in the context of the replication crisis in psychology³⁸) – field of positive psychology, for example mindfulness,³⁹ meditation, and related approaches.

Direct neurotechnologically mediated interventions into the brain intuitively raise health and safety concerns, for example concerning potential adverse effects on mental experience and therefore mental integrity. While such safety concerns are surely reasonable given the direct physical nature of the brain intervention, there is, however, to date no evidence of serious adverse effects for commonly used extracranial electric or electromagnetic neurostimulation techniques such as transcranial direct-current stimulation (tDCS) or repetitive transcranial magnetic stimulation (rTMS).⁴⁰ In stark contrast, comparatively little attention has been paid until recently to the adverse effects of psychological interventions. Studies in the past few years have now demonstrated that seemingly benign interventions such as psychotherapy, mindfulness, or meditation can have discernible and sometimes serious adverse effects on mental health and well-being and thus on mental integrity.⁴¹

³⁵ F Germani and others, ‘Engineering Minds? Ethical Considerations on Biotechnological Approaches to Mental Health, Well-Being, and Human Flourishing’ (*Trends in Biotechnology*, 3 May 2021) <https://doi.org/10.1016/j.tibtech.2021.04.007>; P Kellmeyer, ‘Neurophilosophical and Ethical Aspects of Virtual Reality Therapy in Neurology and Psychiatry’ (2018) 27(4) *Cambridge Quarterly of Healthcare Ethics* 610–627 <https://doi.org/10.1017/S0963180118000129>.

³⁶ CK Kim, A Adhikari, and K Deisseroth, ‘Integration of Optogenetics with Complementary Methodologies in Systems Neuroscience’ (2017) 18(4) *Nature Reviews Neuroscience* 222–235 <https://doi.org/10.1038/nrn.2017.15>.

³⁷ C Janiszewski and RS Wyer, ‘Content and Process Priming: A Review’ (2014) 24(1) *Journal of Consumer Psychology* 96–118 <https://doi.org/10.1016/j.jcps.2013.05.006>; DM Hausman, ‘Nudging and Other Ways of Steering Choices’ (2018) 1 *Intereconomics* 17–20.

³⁸ Open Science Collaboration, ‘Estimating the Reproducibility of Psychological Science’ (2015) 349(6251) *Science* <https://doi.org/10.1126/science.aac4716>.

³⁹ JD Creswell, ‘Mindfulness Interventions’ (2017) 68(1) *Annual Review of Psychology* 491–516 <https://doi.org/10.1146/annurev-psych-042716-051139>.

⁴⁰ H Matsumoto and Y Ugawa, ‘Adverse Events of TDCS and TACS: A Review’ (2017) 2 *Clinical Neurophysiology Practice* 19–25 <https://doi.org/10.1016/j.cnp.2016.12.003>; F Fregni and A Pascual-Leone, ‘Technology Insight: Noninvasive Brain Stimulation in Neurology—Perspectives on the Therapeutic Potential of RTMS and TDCS’ (2007) 3(7) *Nature Clinical Practice Neurology* 383–393 <https://doi.org/10.1038/ncpneu00530>.

⁴¹ AWM Evers and others, ‘Implications of Placebo and Nocebo Effects for Clinical Practice: Expert Consensus’ (2018) 87(4) *Psychotherapy and Psychosomatics* 204–210 <https://doi.org/10.1159/000490354>; WB Britton and others, ‘Defining

Another context in which there is intensive debate around the ethical aspects and societal impact of influencing mental experience and behavior concerns internet-based digital technologies, especially the issue of gamification⁴² and other incentivizing forms of user engagement in ‘social’ media platforms or apps. Certain types of digital behavioral technologies⁴³ are specifically designed to tap into reward-based psychological and neurobiological mechanisms with the aim to maximize user engagement which drives the business model of many companies and developers in the data economy.⁴⁴ While these digital behavioral technologies (DBT) might be used in a healthcare provision context, for example to deliver digital mental health services,⁴⁵ the use of DBT apps in an uncontrolled environment, such as internet-based media and communication platforms raises concern about the long-term impact on mental integrity of users.

To summarize, the quality and content of our mental experience is multifaceted and the ability to successfully integrate different levels of mental experience into a holistic sense of self (as an important component of selfhood or personhood) – mental integrity – is an important prerequisite for mental health and well-being. There are several ways to interfere with mental integrity, through neurotechnologically mediated interventions as well as by many other means. The disruption of the integrated nature of our mental life can lead to severe psychological distress and potentially mental illness. Therefore, protecting our mental life from unwarranted and/or unconsented intervention seems like a justified ethical demand. The law offers many mechanisms for protection in that respect, both at the level of fundamental rights – for example in Article 3 – Right to integrity of the person of the EU Charter of Fundamental Rights⁴⁶ – as well as specific civil laws such as consumer protection laws and medical law.

IV. NEURORIGHTS: LEGAL INNOVATION OR NEW WINE IN LEAKY BOTTLES?

As we have seen in the preceding sections, there are ethically justifiable and scientifically informed reasons to claim that mental privacy and mental integrity are indeed aspects of our human existence (‘anthropological goods’ if you will) that are worthy of being protected by the law. In this section, I will therefore give an overview of recent developments in the legal and

and Measuring Meditation-Related Adverse Effects in Mindfulness-Based Programs’ (*Clinical Psychological Science*, 18 May 2021) <https://doi.org/10.1177/2167702621996340>; M Farias and others, ‘Adverse Events in Meditation Practices and Meditation-Based Therapies: A Systematic Review’ (2020) 142(5) *Acta Psychiatrica Scandinavica* 374–393 <https://doi.org/10.1111/acps.13225>; D Lambert, NH van den Berg, and A Mendrek, ‘Adverse Effects of Meditation: A Review of Observational, Experimental and Case Studies’ (*Current Psychology*, 24 February 2021) <https://doi.org/10.1007/s12144-021-01503-2>.

⁴² A Hoffmann, CA Christmann, and G Bleser, ‘Gamification in Stress Management Apps: A Critical App Review’ (2017) 5(2) *JMIR Serious Games* <https://doi.org/10.2196/games.7216>.

⁴³ L Herzog, P Kellmeyer, and V Wild, ‘Digital Behavioral Technology, Vulnerability and Justice: An Integrated Approach’ (*Review of Social Economy*, 30 June 2021) www.tandfonline.com/doi/full/10.1080/00346764.2021.1943755?scroll=top&needAccess=true (hereafter Herzog, Kellmeyer, and Wild, ‘Digital Behavioral Technology, Vulnerability and Justice: An Integrated Approach’).

⁴⁴ T Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads* (2017); AA Alhassan and others, ‘The Relationship between Addiction to Smartphone Usage and Depression Among Adults: A Cross Sectional Study’ (*BMC Psychiatry*, 25 May 2018) <https://doi.org/10.1186/s12888-018-1745-4>; DT Courtwright, *Age of Addiction: How Bad Habits Became Big Business* (2021); NM Petry and others, ‘An International Consensus for Assessing Internet Gaming Disorder Using the New DSM-5 Approach’ (2014) 109(9) *Addiction* 1399–1406 <https://doi.org/10.1111/add.12457>.

⁴⁵ VW Sze Cheng and others, ‘Gamification in Apps and Technologies for Improving Mental Health and Well-Being: Systematic Review’ (2019) 6(6) *JMIR Mental Health* <https://doi.org/10.2196/13717>.

⁴⁶ EU: Council of the European Union, ‘Charter of Fundamental Rights of the European Union’, C 303/1 2007/C 303/01 § (2007).

policy domain regarding the implementation of such 'neurorights'.⁴⁷ First, I will describe the current debate around the legal foundations and scope of neurorights, then I will propose some conceptual additions to the notion of neurorights and, third, propose a pragmatic and human rights-based approach for making neurorights actionable.

1. *The Current Debate on the Conceptual and Normative Foundations and the Legal Scope of Neurorights*

For a few years now, the debate around the legal foundations and precise scope of neurorights has been steadily growing. From a bird's eye perspective, it seems fair to say that two main positions are dominating the current scholarly discourse: rights conservatism and rights innovationism/reformism. Scholars that argue from a rights conservatism position make the case that the existing set of fundamental rights, as enshrined for example in the Universal Declaration of Human Rights (UDHR) (but also in many constitutional legal frameworks in different states and specific jurisdictions), provides enough coverage to protect the anthropological goods of mental privacy and mental integrity.⁴⁸ Scholars that are arguing from the position of rights innovationism or reformism emphasize that there is something qualitatively special and new about the ways in which emerging neurotechnologies (and other methods, see above) (may) allow for unprecedented access to a person's mental experience or (may) interfere with their mental integrity, and that, therefore, either new fundamental rights are necessary (legal innovation) or existing fundamental rights should be amended or expanded (legal reformism).⁴⁹ Common to both positions is the acknowledgment that the privacy and integrity of mental experience are indeed aspects of human existence that should be protected by the law; the differences in terms of how such mental protection could be implemented, however, have vastly different implications in terms of the consequences for national and international law. Whereas the legal conservatist would have to do the work to show precisely how national, international, and supranational legal frameworks could be effectively applied to protect mental privacy and integrity in specific contexts, the reformist position implies changes in the legal landscape that would have seismic and far-reaching consequences for many areas of the law, national and international policymaking as well as consumer protection and regulatory affairs. From a pragmatic point of view, two major problems immediately present themselves regarding the addition of new fundamental rights that refer to the protection of mental experience to the catalogue of human rights. The first problem concerns the potential for unintended consequences of introducing such novel rights. It is a well-known problem, both in moral philosophy and legal philosophy, that moral and legal goods – especially if they are not conceptually dependent on each other – can (and often do) exist in conflict with each other which, in applied moral philosophy gives rise to classical dilemma situations for example. Therefore, introducing new fundamental rights might serve

⁴⁷ As I am not a legal scholar, this section provides an outside view, informed by my understanding of the neuroscientific facts and ethical discussions, of the current debate at the intersection of neurolaw and neuroethics on the relevance of fundamental rights, particularly international human rights, for protecting mental privacy and mental integrity. In the scholarly debate, this set of issues are usually referred to as 'neurorights' and I will therefore use this term here too.

⁴⁸ S Lighthart and others, 'Forensic Brain-Reading and Mental Privacy in European Human Rights Law: Foundations and Challenges' (*Neuroethics*, 20 June 2020) <https://doi.org/10.1007/s12152-020-09438-4>; C Bublitz, 'Cognitive Liberty or the International Human Right to Freedom of Thought' in J Clausen and N Levy (eds), *Handbook of Neuroethics* (2015) 1309–1333 https://doi.org/10.1007/978-94-007-4707-4_166.

⁴⁹ Yuste and others, 'Four Ethical Priorities for Neurotechnologies and AI' (n 10); Ienca and Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology' (n 10).

the purpose of protecting a specific anthropological good, such as mental privacy, in a granular way, but at the same time it increases the complexity of balancing different fundamental rights and therefore also the potential for moral and/or legal dilemmata situations. Another often voiced criticism is the perceived problem of rights inflation, in other words, the notion that the juridification (German: *Verrechtlichung*) of ethical norms leads to an inflation of fundamental rights – and thus rights-based narratives and juridical claims – that undermine the ability of the polity to effectively address systemic social and other structural injustices.⁵⁰

From my point of view, the current state of this debate suffers from the following two major problems: firstly, an insufficient conceptual specification of mental privacy and mental integrity and, secondly, a lack of transdisciplinary collaborative discourses and proposals for translating the ethical demands that are framed as neurorights into actionable frameworks for responsible and effective governance of neurotechnologies. In the following sections, I address both concerns by suggesting some conceptual additions to the academic framing and discourse around neurorights and proposing a strategy for making neurorights actionable.

2. *New Conceptual Aspects: Mental Privacy and Mental Integrity As Anthropological Goods*

The variability of operational descriptions of mental privacy and mental integrity in the literature shows that both notions are still ‘under construction’ from a conceptual perspective. As important as this ongoing conceptual work is in refining these ideas and for making them accessible to a wide scholarly audience, I would propose here that understanding them mainly as relevant anthropological goods⁵¹ – rather than mainly philosophical or legal concepts – could help to theorize and discuss about mental privacy and mental integrity across disciplinary divides. However, the anthropological goods of mental privacy and mental integrity are conceptually underspecified in the following sense.

First, no clear account is given in the literature of what typical, if not the best approximate, correlates of mental experience (as the substrate of mental privacy) are. Some authors suggest that neurodata or brain data are – or might well become (with advances in neuroscience) – the most direct correlate of mental experience and that, therefore, brain data (and information gleaned from these data) should be considered a noteworthy and special category of personal data.⁵² It could be argued that, in addition to brain data, many different kinds of contextual data (e.g. from smartphones, wearables, digital media and other contexts) allow for similar levels of diagnostic or predictive modelling and inferences on the quality and content of a person’s

⁵⁰ D Clément, ‘Human Rights or Social Justice? The Problem of Rights Inflation’ (2018) 22(2) *The International Journal of Human Rights* 155–169 <https://doi.org/10.1080/13642987.2017.1349245>. Though there are also important objections to these lines of arguments: JT Theilen, ‘The Inflation of Human Rights: A Deconstruction’ (2021) *Leiden Journal of International Law* 1–24 <https://doi.org/10.1017/S0922156521000297>.

⁵¹ An anthropological good, in my usage here, refers to a key foundational dimension of human existence that, throughout history and across cultures, is connected to strong human interests and preferences. Examples would be the interest in and preference for being alive, for having shelter, freedom, food, and so forth. In this understanding, anthropological goods antecede and often are the basis for normative demands, such as ethical claims and rights claims. As a pre-theoretical notion, they are also related to the more developed notion of ‘capabilities’ [M Nussbaum, ‘Capabilities and Social Justice’ (2002) 4(2) *International Studies Review* 123–135 <https://doi.org/10.1111/1521-9488.00258>] insofar as capabilities give a philosophically comprehensive account of how dimensions of human existence relate to fundamental rights.

⁵² Kellmeyer, ‘Big Brain Data’ (n 9); Sara Goering and others, ‘Recommendations for Responsible Development and Application of Neurotechnologies’ (2021) *Neuroethics* <https://doi.org/10.1007/s12152-021-09468-6>.

mental experience.⁵³ What is lacking, however, is a critical discussion of what the right level for protecting a person’s mental privacy is: the level of data protection (data privacy); protecting the information/content that can be extracted from these data (informational privacy); or both; or whether we should also address the question of how and to what ends mental data/information are being used? As discussed above, I would suggest that a very important and legitimate dimension for ethical concerns is also the question of whether and to what extent any kind of neurotechnology or neurodecoding approach has a negative impact on enabling a person to exercise their legitimate interest in their own mental data and information. To be able to respect a person’s interest in data and information on their mental states, however, we would need ethically viable means of disclosing these interests to a third party in ways that do not themselves create additional problems of privacy protection, in other words to avoid a self-perpetuating privacy protection problem. At the level of data and information protection, one strategy could be to establish trustworthy technological means (such as blockchain technology, differential privacy, homomorphic encryption, and other techniques⁵⁴) and/or institutions – data fiduciaries – for handling any data of a person that might allow for inferences on mental experience.

Second, the demand for protecting mental integrity is undermined by the problem that we do not have a consensual conceptual understanding of key notions such as agency, autonomy, and the self. Take the example of psychedelic recreational drugs, as an example for an outside interference with mental integrity. We have ample evidence from psychological and psychiatric research that suggests that certain types of recreational psychedelic drugs, such as LSD or Psilocybin, have discernible effects on mental experiences associated with personal identity and self-experience, variously called, for example, ‘ego dissolution’⁵⁵ or ‘boundlessness’.⁵⁶ However, most systematic research studying these effects, say in experimental psychology or psychiatry, is not predicated on a universal understanding or model of human self-experience, personal identity, and related notions. As even any preliminary engagement with conceptual models of personal identity or ‘the’ self in psychology, cognitive science, and philosophy will quickly reveal, there are indeed many different competing, often conceptually non-overlapping or incommensurable models available: ranging from constructivist ideas of a ‘narrative self’, to embodiment-related (or more generally 4E-cognition-related) notions of an ‘embodied self’ or ‘active self’, to more socially inspired notions such as the ‘relational self’ or ‘social self’.⁵⁷

⁵³ Herzog, Kellmeyer, and Wild, ‘Digital Behavioral Technology, Vulnerability and Justice: An Integrated Approach’ (n 42); KV Kreitmair, MK Cho, and DC Magnus, ‘Consent and Engagement, Security, and Authentic Living Using Wearable and Mobile Health Technology’ (2017) 35(7) *Nature Biotechnology* 617–620 <https://doi.org/10.1038/nbt.3887>; N Minielly, V Hrinco, and J Illes, ‘A View on Incidental Findings and Adverse Events Associated with Neurowearables in the Consumer Marketplace’ in I Bárd and E Hildt (eds), *Developments in Neuroethics and Bioethics*, vol. 3 (2020) 267–277 <https://doi.org/10.1016/bs.dnb.2020.03.010>.

⁵⁴ V Jaiman and V Urovi, ‘A Consent Model for Blockchain-Based Health Data Sharing Platforms’ in *IEEE Access* 8 (2020) 143734–143745 <https://doi.org/10.1109/ACCESS.2020.3014565>; A Khedr and G Gulak, ‘SecureMed: Secure Medical Computation Using GPU-Accelerated Homomorphic Encryption Scheme’ (2018) 22(2) *IEEE Journal of Biomedical and Health Informatics* 597–606 <https://doi.org/10.1109/JBHI.2017.2657458>; MU Hassan, MH Rehmani, and J Chen, ‘Differential Privacy Techniques for Cyber Physical Systems: A Survey’ (2020) 22(1) *IEEE Communications Surveys Tutorials* 746–789 <https://doi.org/10.1109/COMST.2019.2944748>.

⁵⁵ C Letheby and P Gerrans, ‘Self Unbound: Ego Dissolution in Psychedelic Experience’ (2017) 1 *Neuroscience of Consciousness* <https://doi.org/10.1093/nc/nix016>.

⁵⁶ FX Vollenweider and KH Preller, ‘Psychedelic Drugs: Neurobiology and Potential for Treatment of Psychiatric Disorders’ (2020) 21(11) *Nature Reviews Neuroscience* 611–624 <https://doi.org/10.1038/s41583-020-0367-2>.

⁵⁷ PT Durbin, ‘Brain Research and the Social Self in a Technological Culture’ (2017) 32(2) *AI & SOCIETY* 253–260 <https://doi.org/10.1007/s00146-015-0609-4>; S Gallagher, ‘A Pattern Theory of Self’ (2013) 7 *Frontiers in Human Neuroscience* <https://doi.org/10.3389/fnhum.2013.00443>; T Fuchs, *The Embodied Self: Dimensions, Coherence, and Disorders* (2010); D Parfit, ‘Personal Identity’ (1971) 80(1) *The Philosophical Review* 3–27.

Consequently, any interpretation, let alone systematic understanding, of how certain interventions might or might not affect mental integrity – here represented by the dimension of self-experience and personal identity – will heavily depend on the conceptual model of mental experience that one has. This rather obvious point about the inevitable interdependencies between theory-driven modelling and data-driven inferences and interpretation has important consequences for the ethical demands and rights-claims that characterize the debate on the neurorights. First, this should lead to the demand and recommendation that any empirical research that investigates the relationship between physical (for instance via neurotechnologies or drugs) or psychological interventions (for example through behavioural psychology, such as nudging) and mental experience should make their underlying model of self-experience and personal identity explicit and specify it in a conceptually rigorous manner. Second, transdisciplinary research on the conceptual foundations of mental (self-)experience, involving philosophers, cognitive scientists, psychologists, neuroscientists, and clinicians should be encouraged to arrive at more widely accepted working models that can then be tested empirically.

3. *Making Neurorights Actionable and Justiciable: A Human Rights–Based Approach*

Irrespective of whether new fundamental rights will ultimately be deemed necessary or whether existing fundamental rights will prove sufficient to protect the anthropological goods mental privacy and mental integrity, regulation and governance of complex emerging sciences and technologies, such as AI-based neurotechnology, is a daunting challenge. If one would agree that reasonable demands for any governance regime that allows innovation of emerging technologies in a responsible manner include that the regime is context-sensitive, adaptive, anticipatory, effective, agile, and at the right level of ethical and legal granularity, then the scattered and inhomogeneous landscape of national and international regulatory and legal frameworks and instruments presents a particularly complex problem of technology governance.⁵⁸

Apart from the conceptual issues discussed here that need to be further clarified to elucidate the basis for specific ethical/normative demands for protecting mental privacy and mental integrity, another important step for making neurorights actionable is finding the right levels of governance and regulation and appropriate (and proportional) granularities of legal frameworks. So far, no multi-level approach to legal protection of mental privacy and mental integrity is available. Instead, we find various proposals and initiatives at different levels: at the level of ethical self-regulation and self-governance; represented for example by ethical codes of conduct in the context of neuroscience research⁵⁹ or in the private sector around AI governance;⁶⁰ at the level of national policy, regulatory, and legislative initiatives (e.g. in Chile);⁶¹ at the level of supranational policies and treaties, represented, for example, by the intergovernmental report on

⁵⁸ More generally the complexity of the legal landscape and political processes creates the well-known ‘pacing problem’ in governing and regulating technological innovations, also referred to as the ‘Collingridge Dilemma’, cf. for example: A Genus and A Stirling, ‘Collingridge and the Dilemma of Control: Towards Responsible and Accountable Innovation’ (2018) 47(1) *Research Policy* 61–69 <https://doi.org/10.1016/j.respol.2017.09.012>.

⁵⁹ Exemplified by the Ethics Policy of the Society for Neuroscience, the largest professional body representing neuroscience researchers: SfN, ‘Professional Conduct’ (SfN) <https://www.sfn.org/about/professional-conduct>.

⁶⁰ Consider for example: *Partnership on AI* www.partnershiponai.org/.

⁶¹ L Dayton, ‘Call for Human Rights Protections on Emerging Brain-Computer Interface Technologies’ (*Nature Index*, 16 March 2021) <https://www.natureindex.com/news-blog/human-rights-protections-artificial-intelligence-neurorights-brain-computer-interface>.

responsible innovation in neurotechnology of the Organization for Economic Co-operation and Development (OECD) from 2019⁶².

Taking these complex problems into account, I would advocate for a pragmatic, human rights-based approach to regulating and governing AI-based neurotechnologies and for protecting mental privacy and mental integrity as anthropological goods. This approach is predicated on the assumption that existing fundamental rights, as enshrined in the UDHR and many national constitutional laws, such as the right to freedom of thought,⁶³ provide sufficient normative foundations. On top of these foundations, however, a multi-level governance approach is required that provides context-sensitive and adaptive regulatory, legal, and political solutions (at the right level of granularity) for protecting humans from potential threats to mental privacy and mental integrity, such as in the context of hitherto un- or underregulated consumer neurotechnologies. Such a complex web of legal and governance tools will likely include bottom-up instruments, such as ethical self-regulation, but also laws (constitutional laws, but also consumer protection laws and other civil laws) and regulations (data protection regulations and consumer regulations) at the national level and supranational level, as well as soft-law instruments at the supranational level (such as the OECD framework for responsible innovation of neurotechnology, or widely adopted ethics declarations from specialized agencies of the United Nations (UN), such as UN Educational, Scientific and Cultural Organization (UNESCO) or World Health Organization (WHO)).

But making any fundamental right actionable (and justiciable) at all levels of societies and international communities requires a legally binding and ethically weighty framework to resolve current, complex, and controversial issues in science, society, and science policy. Therefore, conceptualizing neurorights as a scientifically grounded and normatively oriented bundle of fundamental rights (and applied legal and political translational mechanisms) may have substantial inspirational and instrumental value for ensuring that the innovation potential of neurotechnologies, especially AI-based approaches, can be leveraged for applications that promote human health, well-being, and flourishing.

V. SUMMARY AND CONCLUSIONS

In summary, neurorights have become an important subject for scholarly debate, driven partly by innovation in AI-based decoding of neural activity, and as a result different positions are emerging in the discussion around the legal status of brain data and the legal approach to protecting the brain and mental content from unwarranted access and interference.

I have argued that mental privacy and mental integrity could be understood as important anthropological goods that need to be protected from unwarranted and undue interference, for example, by means of neurotechnology, particularly AI-based neurotechnology.

In the debate on the question of how neurorights relate to existing national and supranational legal frameworks, especially to human rights, three distinct positions are emerging: (a) a rights conservatism position, in which scholars argue that existing fundamental rights (e.g. constitutional rights at the national level and human rights at the supranational level) provide adequate protection to mental privacy and mental integrity; (b) a reformist, innovationist position, in which scholars argue that existing legal frameworks are not sufficient to protect

⁶² OECD Legal Documents, 'Recommendation of the Council on Responsible Innovation in Neurotechnology' <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457>.

⁶³ Article 18, UDHR.

the brain and mental content of individuals under envisioned near-future scenarios of AI-based brain decoding through neurotechnologies and, therefore, reforms of existing frameworks – such as constitutional laws or even the Universal Declaration of Human Rights – are required; and (c) a human rights-based approach, that acknowledges that the law (in most national jurisdictions as well as internationally) provides sufficient legal instruments but that its scattered nature – across jurisdictions as well as different areas and levels of the law (such as consumer protection laws, constitutional rights, etc.) – requires an approach that makes neurorights actionable and justiciable, for example by connecting fundamental rights to specific applied laws (e.g. in consumer protection laws).

The latter position – which in the policy domain would translate into a multi-level governance approach – has the advantage that it does not argue from entrenched positions with little room for consilience but provides deliberative space in which agreements, treaties, soft law declarations, and similar instruments for supra- and transnational harmonization can thrive.