Neuroethics and AI ethics: a proposal for collaboration

Arleen Salles[1]* and Michele Farisco[2,3]*

Abstract

The scientific relationship between neuroscience and artificial intelligence is generally acknowledged, and the role that their long history of collaboration has played in advancing both fields is often emphasized. Beyond the important scientific insights provided by their collaborative development, both neuroscience and AI raise a number of ethical issues that are generally explored by neuroethics and AI ethics. Neuroethics and AI ethics have been gaining prominence in the last few decades, and they are typically carried out by different research communities. However, considering the evolving landscape of AI-assisted neurotechnologies and the various conceptual and practical intersections between AI and neuroscience—such as the increasing application of AI in neuroscientific research, the healthcare of neurological and mental diseases, and the use of neuroscientific knowledge as inspiration for AI—some scholars are now calling for a collaborative relationship between these two domains. This article seeks to explore how a collaborative relationship between neuroethics and AI ethics can stimulate theoretical and, ideally, governance efforts. First, we offer some reasons for calling for the collaboration of the ethical reflection on neuroscientific innovations and AI. Next, we explore some dimensions that we think could be enhanced by the cross-fertilization between these two subfields of ethics. We believe that considering the pace and increasing fusion of neuroscience and AI in the development of innovations, broad and underspecified calls for responsibility that do not consider insights from different ethics subfields will only be partially successful in promoting meaningful changes in both research and applications.

Keywords Neuroethics, AI ethics, Culture, Responsible conceptualization, Governance

Background

The scientific relationship between neuroscience and artificial intelligence (AI) is generally acknowledged, and the role that their long history of collaboration has played in advancing both fields is often emphasized [29, 35, 91]. Beyond the important scientific insights provided

*Correspondence:
Arleen Salles
asalles@instituteofneuroethics.org
Michele Farisco
michele.farisco@uu.se
[1] Institute of Neuroethics (IoNx), Atlanta, GA, USA
[2] Centre for Research Ethics and Bioethics, Uppsala University, Uppsala, Sweden
[3] Biogem, Biology and Molecular Genetics Research Institute, Bioethics Unit, Ariano Irpino, AV, Italy

by their collaborative development, both neuroscience and AI raise a number of ethical issues that are generally explored by neuroethics and the ethics of AI usually cultivated as two separate subfields of ethics.

Neuroethics can be broadly defined as a field that addresses philosophical, ethical, legal, social, and cultural questions raised by neuroscience and related technologies [27, 42, 45, 51, 60]. It emerged as a response to advances in neuroscience that have been challenging ingrained notions and related ethical views about several topics, including the structure and function of the peripheral and central nervous system, the basis of consciousness, the brain-mind relationship, and the foundations of our humanness itself.

AI ethics refers to the area of inquiry that addresses the social, regulatory, ethical, and philosophical dimensions

of the development and use of AI [28]. The field attempts to formulate and develop theoretical and practical approaches to anticipate and minimize the potential adverse effects of AI research and applications across a diverse range of social and economic activities, and to enhance the advantages of AI for society.

Neuroethics and AI ethics have been gaining prominence in the last few decades, and they are typically carried out by different research communities. However, in light of evolving AI- assisted neurotechnologies and other conceptual and practical intersections between AI and neuroscience, such as the increasing application of AI in neuroscientific research, the treatment of neurological and mental diseases, and the use of neuroscientific knowledge as inspiration for AI- some scholars are now calling for a collaborative relationship between these two areas of research and practice.

This article aims to build from this proposal, seeking to further explore how a collaborative relationship between neuroethics and AI ethics can stimulate theoretical and, ideally, governance efforts. First, we begin by reviewing some reasons for calling for the collaboration of neuroethics and AI ethics. We then explore the dimensions that we think could be enhanced by the cross-fertilization between them. We believe that considering the pace and increasing fusion of neuroscience and AI in the development of innovations, broad and underspecified calls for responsibility that do not consider insights from different ethics subfields will only be partially successful in promoting meaningful changes in both research and applications.

Main text

Why collaboration?

While neuroethics and AI ethics have developed independently from one another, recently there have been calls for a collaborative discussion of the issues addressed by these subfields of ethics [6, 8, 26, 40, 41, 49].[1] The need for such collaboration is grounded on the recognition of significant commonalities within the fields of neuroscience and AI: specifically, overlapping domains of research and application (i.e., shared contents), common use of fundamental concepts (i.e., shared categories), and

some common fundamental concerns and challenges (i.e., shared drivers and aims).

a. Overlapping domains: AI technology is increasingly assisting in several brain-related areas [93]. It is used not just to enhance neuroscientific research generating new knowledge (e.g., through optimized handling of medical data revealing new connections among them and informing more efficient predictions) but in practice for early diagnosis in brain and mental health, to improve the design and efficiency of existing neurotechnologies (e.g., Brain–Computer Interfaces (BCI)), in wearable devices for monitoring and screening brain activity, and to tailor existing drugs to the needs of individual patients, among others [69], as well as to generate new powerful bioorganic computer chips which use the computational power of the human brain, more specifically of brain organoids [13, 88].

b. Common use of concepts: on one hand, AI researchers often use ontological, psychological, and normative notions and terms borrowed from neuroscience, even if occasionally these are adapted and even re-conceptualized (e.g., intelligence, consciousness, learning, neurons, synapses, among others). On the other hand, the computational metaphor about the brain (i.e., its view as an information processing device) even if less popular than in the second half of the last century [19] is still used to describe it. To illustrate, the current discussion about the possibility of conscious AI often relies on computational functionalism [12], which depicts consciousness as the result of the right computation in the right physical structure, and the brain as a particular computational system that may be replicated in AI systems.[2]

c. Common or similar ethico-societal issues at the practical and theoretical levels:

   a. AI and neurotechnology applications often raise questions of bias and stigma as well as concerns about their potential impact on privacy, decision making, the workplace, dual use, and human rights, among others [6, 21, 32]. Some of these risks can actually be increased by the combination of advances in neuroscience and AI.

   b. AI and neurotechnology raise issues about potentially transforming the *status quo* triggering

---

[1] Of course, neuroethics is not the only ethics subfield that can productively collaborate with AI ethics. Computing ethics, digital ethics, information ethics, and machine ethics, among others, have bodies of work addressing ethical issues relevant to the design and deployment of AI. Importantly, however, the conceptual boundaries between Ai ethics and these other technology ethics subfields might be blurred whereas this is not the case with neuroethics which is not typically considered a type of technology ethics. For a rich discussion on the proliferation of diverse technology ethics subfields see Saetra and Danaher [77] and Llorca-Albareda and Rueda [54].

[2] This view has been criticized by some researchers in neuroscience, who stress that even if the brain may be conceived as a computational system, it is organized in multiple levels and scales, which confer a high level of complexity (i.e., dynamic interaction between several components) to it [1, 90]

uncertainty about society and the world in the future [58].

c. Underlying many of the issues raised by the application of AI and of some neurotechnologies are philosophical convictions about the nature of humanness itself, about the line between humans and machines, and the fear of potential threats to human agency, autonomy, and dignity [8, 18, 83, 84]. This common conceptual background often results in similar drivers and aims for both neuroscience and AI as well as for the ethical reflection about them.

In view of these commonalities, bringing together the insights and developments from neuroethics and AI ethics is promising. Their collaboration can make the ethical reflection more effective (e.g., in identifying and anticipating emerging issues and in elaborating concrete and effective solutions). Indeed, below we suggest that the existing compartmentalization in addressing the issues has a negative impact on the identification and discussion of some key topics of concern. We present a few areas where collaboration promises to enrich the analysis and management of the relevant issues.

Responsible conceptualization

While calls for responsible research and innovation have been prevalent and generally discussed in the last few years, responsible conceptualization has not been typically recognized as one of its key elements.

We characterize responsible conceptualization as the process of improving conceptual clarification in science and technology with the specific aim to enhance not just scientific research and technological development but also the ethical and normative discussion of the issues they raise as well as the determination of how to address them.[3] The underlying idea is that clear concepts are not only a prerequisite for moving science and innovation forward: they are also key for doing so responsibly (i.e., aligning social needs and preferences with the research and innovation agenda so as to truly serve public good). Conceptual clarity enhances the identification, understanding, and management of the ethical and social issues

raised by research and innovation and improves scientific communication with diverse publics.

Indeed, how concepts themselves are characterized is not only theoretically important, it is key in the communication of meanings and the narratives used to discuss science, its significance, and its outputs, and plays a big role in determining ethical priorities, citizens' attitudes towards science and innovations, and financial support for research and innovations [4, 23, 67, 80].

Awareness of this fact is very relevant in fields like AI and neuroscience. Consider, for example, the field of AI: it uses many familiar concepts (e.g., intelligence, consciousness, autonomy, agency, learning, training, goal, reward) that because of their familiarity might seem to require no further clarification. And yet, when trying to define them, the lack of consensus regarding both what they mean, and their contextual appropriateness becomes clear. While conceptual ambiguity generally does not raise significant difficulties in day-to-day activities, vagueness may be problematic when we are trying to identify what are the societal and ethical concerns raised by AI, and how to address them productively. In fact, some of the ethical questions about harms, benefits, responsibility typically raised when discussing AI are often grounded on not fully accurate understanding of what it is and does [44].

Take, for instance, some fears and concerns raised by the possibility of conscious AI. The fact is that the conceivability of conscious AI depends on how consciousness (a controversial notion indeed) is understood in the first place [26]. If conceived as a biological phenomenon for which the biological component plays a crucial role [70, 86], then reproducing consciousness in a non-biological system would not be achievable, and ethical concerns regarding conscious AI would thus be groundless. However, if consciousness is understood within a functionalist and computational framework, its artificial implementation is conceptually consistent even if still practically not possible [20]. We would have a similar case if consciousness were conceived as a multidimensional feature, where it would be possible to artificially replicate selected dimensions [22]. These last two cases would make ethical concerns about conscious AI possibly premature and misleading, but not misplaced.

Now let's consider an initially less controversial notion: learning. Learning is a term widely used to describe the training of artificial systems, especially Deep Learning (DL) systems (e.g., LLMs). This notion, however, becomes problematic because it may lead people to attribute other directly related features (e.g., experience, competence, flexibility, and even wisdom) to AI, features that are currently hardly applicable to AI systems. Even if there is a basic analogy between what happens in the human brain

---

[3] We do not engage here with an important and ultimately complementary issue that we will be addressing in a different paper: the role of conceptual engineering as a response to the conceptual disruption caused by emerging technologies (see, for example, [37, 59]). The responsible conceptualization that we propose might be seen as a component of conceptual engineering understood as conceptual adaptation (or even amelioration). In fact, responsible conceptualization proposes to adapt and/or improve our moral concepts in order to better face the specific ethical challenges arising from new technologies.

and in DL architecture when they "learn" (i.e., a change in the inter-neurons connections or weights), the two concepts do not really overlap. The fundamental difference is the level of generalization and flexibility that human learners eventually display, which goes far beyond the limited competence of DL systems, which are dependent on their training data. This fundamental difference connects to several other disanalogies, including the capacity of humans for creativity, which allows them to solve new kinds of problems they have never faced before [16].

Conceptual issues about consciousness and sentience are often also present in the field of neuroscience and in the description of some of its outputs: consider neural organoid research. The issue of consciousness, including the moral status of organoids, is a key point of discussion [38, 66, 81]. In turn, the recent controversy over the use of the term "sentience" to refer to cultures of human and mouse neurons that exhibit goal-directed activity adaptations [3, 47, 75] revolved partly around the role of conceptual clarity in interpreting research findings and the problem of ambiguous or misleading conceptualization and language in scientific progress and in the discussion of the ethical and social issues that science raises. In this specific case, the debate emerges because the term "sentience," which is inherently ambiguous [65], is often used to denote a form of consciousness that people generally regard as morally significant [52]. The need for terminological and conceptual caution in using this term is equally important in both neuroscience and AI (e.g., in the discussion about synthetic biological intelligence or artificial consciousness) [48, 85].

Technical terms can also be conceptually vague or misleading. Consider the term "digital twin" to refer to brain network models created to support diagnostic and therapeutic interventions [23, 53]. Computational brain models meet some of the basic conditions for digital twinness (i.e., a seamless connection with the physical entity that they are simulating or mirroring, a real-time data exchange between model and brain) so initially the adequacy of using the term "digital twin" might seem uncontroversial. However, ethically relevant conceptual issues still arise: computational models of the brain are not digital replicas of brains; instead they replicate very specific and targeted brain functions. As a result, their level of fidelity to their physical counterpart is more limited than the term might suggest to non-expert publics. However, if non-expert publics are aware of the creation of digital twins of the brain, they are likely to infer the existence of a type of entity that is different from what these models actually represent [23]. This is one reason for finding the use of terms such as "digital twin" to refer to these models problematic. Similar points have been made regarding the use of the term "mini brain" to refer

to brain organoids [5]. Indeed, the issue of determining what level of functional and structural fidelity is required for computational models (which includes AI systems based on such models) to be considered a reliable simulation of the target object (either neural or non neural) is key and one that is shared by AI and brain research. This supports the need for collaboration when addressing this concern and the relevant ethical and social impacts.

The lack of conceptual clarity can have clear practical implications: insofar as understandings shape people's attitudes towards scientific outcomes, the use of unclear concepts increases the likelihood of confusion, mistaken beliefs, and false expectations both within and beyond the scientific communities [4, 23].

In the last few years, the importance of attending to conceptual clarity and language in neuroscientific practices and of recognizing its role in promoting good scientific practices has been receiving some attention within the field of neuroethics. There have been concerted efforts to unveil and address conceptual issues [14, 15, 24, 27, 61, 74, 79, 87] and to identify the implications of the lack of conceptual clarity at the theoretical and practical levels. Because conceptual clarity impacts both a research lifecycle and its future applications, some neuroethicists call for using conceptual analysis to provide a semantic clarification of the relevant terms including their use and epistemic and ontological adequacy in diverse contexts [24, 27, 79]. The idea is that conceptual clarification is not just an important tool to better neuroscientific practice itself, but key to inform, refine, and enhance the ethical analysis of the issues raised by neuroscientific research and applications.

Conceptual efforts to enhance both scientific practices themselves and the discussion of the ethical issues they raise could be further advanced by collaboration between neuroethicists and AI ethicists. This could be done by, for example, doing joint work on specific concepts (e.g., intelligence, decision-making, consciousness, autonomy) that play a key role in both fields. This joint work would include reconsidering their foundation and unveiling underlying assumptions, focusing on how each field conceptualizes each notion and why, making the different conceptualizations visible, and exploring their theoretical and practical implications and ways to bridge them. This would benefit not just those who do the research (by shedding light on the epistemic justification of the concepts they use) but also non-expert publics (by enhancing their understanding of each of these disciplines and their synergies, and their capacity to assess what are the issues to address and why).

The responsible conceptualization that we have in mind requires that concepts and terminology be seen as key to scientific practice, to the interpretation of scientific

findings, to the clarification of their ethical relevance and salience, and to an assessment of their desirability (e.g., anticipating their social impact). In turn, collaboration in responsible conceptualization means that conceptual clarification of some shared terms in neuroscientific and AI innovations should not be siloed and carried out in different spaces: an encompassing ethical analysis requires an overarching definition of those concepts, beyond disciplinary boundaries. Use of the relevant terms should reflect awareness of interpretations in each of the fields and awareness of how each field is using the terms, as well as awareness of possible commonalities in defining and using those terms.

Cultural diversity

Above we proposed that responsible conceptualization will benefit from and often requires the joint conceptual work of neuroethics and AI ethics, especially for those ethically relevant concepts which are used by both fields and are therefore exposed to the risk of inconsistent understandings. Now we turn to another issue that often inadvertently shapes the identification and discussion of the ethical issues raised by innovations: cultural diversity.

For the purposes of this article, we understand culture as information (beliefs, values, assumptions) transmitted from one individual and/or group to another with an overt or covert impact on their thinking and behavior. Cultures can be demarcated by socio-anthropological factors (e.g., ethnicity, race, geographical regions) or by disciplinary or even organizational factors that often shape the understanding and perspectives of those who operate within disciplines/organizations. Here, our concern is with cultural diversity related to anthropological and sociological dimensions.

There are at least three facts that make the call for attention to culture in neuroscience and AI research and outputs compelling. First, both neuroscience and AI research and related innovations can have a significant impact on culture (including socio-political values) at the local, international, and global levels, providing new forms of knowledge, ways to interact with each other, and to think about humans and societies in general. Second, political and other non-epistemic values influence neuroscience and AI research and development, having an impact on scientific research priorities and funding, on which goals are pursued, and which technological outputs should be commercialized and could have an impact on society. Third, cultural contexts can shape how the terms used in research and the development of innovations are interpreted, as well as how the ethical and societal issues they raise are framed and addressed. In short, there is a reciprocal influence between cultural diversity on the one hand and neuroscience and AI on the other hand.

The above-mentioned three facts are ethically relevant. Regarding the first, advances in neurotechnology and AI have opened new horizons to many people, but not all. They have also raised concerns about whether the increased availability of neuro and AI innovations might increase the vulnerability of some groups, and exacerbate regional and global disparities (e.g., in terms of knowledge of and access to new technologies). The second fact appears implicitly and, sometimes, explicitly in the discussion over potential and actual bias in the design, development and deployment of AI. Within AI ethics and neuroethics the first and second aspects have been addressed in several articles and books [2, 18, 31, 39, 68–73]. Here we focus on the third which has received less attention.

What does it mean to say that cultural factors shape interpretations, framings, and conceptualization? Consider the brain: people's understanding of what it is and does has morphed throughout history, shaped by specific historical and cultural contexts. Importantly, available technologies have played an important role in shaping the metaphors that have been used to describe and give meaning to it (e.g., in the past, the brain was explained referring to the telegraph or telephone, the most advanced technologies then available, while in more recent time we have tended to describe the brain as a computer) [17]. The same is true of the notion of mind, which underwent a long process of naturalization and whose conceptualization has been impacted by historically contingent cultural models, both from science and religion [55]. We also see cultural influence in the conceptualization of, for example, mental illness [50, 82], as well as in notions that contribute to shaping people's view of AI, both as professionals in the field and as lay people [34].

Additionally, culture influences which scientific outputs are adopted and how they are integrated in diverse societies as well as decisions about which topics should be prioritized in ethical reflection. Indeed, lack of cultural awareness can thwart collaboration, limit the sharing of the results of neuroscientific findings and hinder awareness of the short- and long-term potential and risks of neuroscientific research (Global Neuroethics Summit [30]). And yet, some have noted that neuroethics, despite increasing attention to diversity, is rooted in Western culture and predominantly focuses on issues that affect users of new neurotechnology. This focus, however, does not adequately represent globally significant issues or alternative approaches to neurological and psychiatric health. Issues of social inequality and their impact on neurological and psychiatric health receive minimal

attention and are typically not a fundamental feature of neuroethics scholarship or bioethics in general [68]. This strongly suggests that culture is a driver of ethical scrutiny by shaping what is deemed ethically problematic and determining which issues deserve ethical reflection.

The fact that interpretations, conceptualizations, and framings are informed by cultural considerations is ethically relevant not only because, as argued in the section above, they play a key role in the ethical discussion but also because they naturally influence debates on governance. Interestingly, ongoing international discussions about the governance of AI and neurotechnologies often seem to assume a culturally uniform perspective on key concepts, ethical issues, and possible solutions. There are good reasons for doing this: as the use of these innovations increases, achieving some level of alignment in ethical principles or at least agreeing on a foundational ethical framework to manage some of the questions and concerns is seen as desirable [94]. However, given the lack of cultural diversity, it is important to recognize two things. First, aspirations to universality may inadvertently mask cultural dimensions that might impact people's understanding of the ethical issues and affect the operationalization and, ultimately, the success of proposed governance frameworks [11, 36, 78]. Second, an over emphasis on consensus of values might unintentionally lead to overlooking ethically relevant cultural contextualities and, in the worst case, might be perceived by some communities as a type of "colonialism." Therefore, it is important to promote awareness of and sensitivity to cultural diversity and to elaborate concrete strategies for respecting such diversity. This should be done without falling into oversimplification, stereotyping, hyperbole, homologation, marginalization, idealization, trivialization, or relativism. Unsurprisingly, how to resolve the tension between the need for recognizing and attending to cultural diversity and the need for some type of global governance remains an open question, but this should not lead us to downplay either need.

Within neuroethics, a significant step toward resolving this tension includes recent proposals for a meaningful engagement with diverse publics and joint reflection on neuroethics questions (Global Neuroethics Summit [30]) across cultures. In particular, Karen Rommelfanger and Laura Specker Sullivan argue for a robust cross cultural approach that seeks to identify similarities against a background of differences, or to highlight differences against a background of similarities where the focus is put on relation [76]. In the authors' view, this type of cross-cultural work can be carried out in different ways, from participation in multicultural meetings to research and capacity building. Importantly, we would argue that insofar as the goal is to foster intercultural understanding, identify

shared concerns, enhance intra-cultural creativity, and at the same time enable a deeper understanding of the distinctive aspects of one's culture while avoiding essentializing any one culture, attention to diverse culturally shaped conceptualizations of some of the main notions remains key.

It is true that how ethical and cultural issues might be manifested in AI ethics might be in some respects different from how they are manifested within neuroethics. To illustrate, the global landscape in neurotechnology shows that even if countries such as South Korea and China are showing an increase in development and patenting, the United States has historically led the field, ranking higher than other areas in scientific publications, investment, and patent applications [33]. Additionally, the main neurotechnology companies are located in the US and Europe (https://www.neurotech.com/charts). This would explain (although not excuse) the westernized tone of the ethical discussion. In contrast, in the context of AI research and innovation, investment is more evenly distributed across several countries in diverse regions, notably, the United States, China, United Kingdom, India, and Germany among others (https://ifamagazine.com/about/)). Accordingly, North America, Europe, and East Asia shape how the main notions are conceptualized and discussed as well as the debate over the ethics and governance of AI [64]. Still, this does not mean that all the issues are equally conceived in all places, nor that they resonate in the same way. A cross-cultural approach such as the one described is expected to facilitate a richer understanding of the cultures involved and a bridging of the dichotomous thinking often present in AI discourse frequently driven by a rhetoric that frames AI development as the next space race.

The convergence of neuroscience, neurotechnology, and AI, and the deployment of AI-assisted neurotechnologies in different parts of the world makes identification and understanding of cultural assumptions and framings particularly relevant. How these technologies are understood (e.g., their technical and social utility), which meaning people assign to them, and how they assess their societal adequacy is often not the same for every culture. In short, the place of these innovations in different societies is not solely dependent on the quality of the innovation, but shaped by local meanings, institutions, and structures. The process of identification and examination of the role that cultural contextualities play in the design and deployment of neurotechnological and AI innovations benefits from breaking down siloed approaches.

Governance
In general, the discussion over AI governance has tended to be independent from the discussion over governance

for neurotechnology. Policy-makers, academics, and several members of the AI research and user community have shown interest in addressing the governance of AI at different levels, as shown by the proliferation of both academic papers [92] and numerous AI ethics standards and recommendations [43] from international general instruments such as the OECD Recommendation of the Council on Artificial Intelligence, and the UNESCO Recommendation on the Ethics of Artificial Intelligence to codes of ethics produced by professional bodies for their members and practitioners or by other regulatory bodies such as the (IEEE Code of Ethics) as well as by recent efforts to pass laws to regulate it (The EU Artificial Intelligence Act). The existence of guidelines, recommendations, and specific attempts at regulation manifest awareness of potential and actual issues raised by AI, and some proactivity in reflecting on them and addressing how to manage them, as well as productive discussions about the challenges of implementation and of translating ethical guidelines into actionable strategies [89].

There are ethics guidance documents within the neuroscience and neurotech communities as well [6, 56, 62]. Importantly, the rapid development of diverse neurotechnologies, their wider applicability, and the recognition that the economic, social, and ethical impacts of their deployment can scale easily and outpace ethical and societal reflection [95] have led to recent calls for the development of international regulatory instruments [11, 63].

A variety of concerns have emerged in the debate over the governance of emerging technologies in general not least the issue of how to fill the gap between development and legislation [57], what kind of regulatory model to use [10, 58, 96] and which role the state should play in governance mechanisms [9]. A number of factors have been identified as important to consider in the discussion over the regulation of neurotechnology, from the issue of the philosophical, legal, and normative basis of any regulatory instrument intended to manage actual and potential risks and the need for consensus on what the goals of such regulatory instrument should be [11], to more specific concerns such as the implications of the spillover of medical neurotechnology into the consumer market, and the nature and status of neural data, its relation with mental data, and whether it requires special attention and protection.

As noted earlier, AI and neurotechnology innovations often raise many similar practical issues (e.g., autonomy, bias) and tend to generate similar philosophical concerns (e.g., identity, relation between humans and machines) even if those concerns are not identical. Let's consider the cases of autonomy and bias. Concerning autonomy, it has been argued that it

poses significant challenges in neurotechnology for a number of reasons, including that neurotechnological devices have the potential to directly access and collect neurodata and even modulate or stimulate the nervous systems, often without detection [7]. AI also raises related issues, such as relying on vast amounts of data often collected without explicit consent and possessing an impressive capacity to process that data. This capability allows AI to detect hidden patterns and access sensitive personal information that may be exploited for various purposes without the data subject's explicit consent.

Concerning bias, its presence in the development, discovery, and interpretation of neurotechnology is widely acknowledged [32]. Neurotechnologies are often based on analysis of datasets from homogeneous populations and training samples which can skew research goals, interpretations, and assessment, while possibly leading to exclusion or misrepresentation of minority and vulnerable populations. Moreover, biased data affects what is considered "normal" brain function and the ethical implications of neuro technological advancements [7]. In turn, the prevalence of algorithmic bias within AI is also a topic of concern. A major cause of algorithmic bias is the data used to train algorithms. If historical data includes biases related to gender, race, or other factors, the algorithm can learn and continue these biases. Additionally, biases can also emerge during data collection, for example, when certain groups are underrepresented in the data, the algorithm may not perform accurately for those groups [25].

The integration of neuroscience, big data, and AI—allowing for retrospective, real-time, and predictive exploration of connections between data patterns and specific mental activities—is enhancing the development and expanding the application of various neurotechnologies (e.g., BCIs) [46]. Moreover, this convergence is intrinsically tied to the functionality of these devices. Promising as this convergence is, however, the evolving nature of both AI and neurotechnological innovations, their growing interdependence, and the continued attempt to extend their domain of application is likely to increase the emergence of some practical issues—i.e. safety, security, dual use, and privacy related issues and uncertainties- and foundational (or fundamental) issues—i.e. issues that might impact our understanding of fundamental notions (e.g., personhood, autonomy) and even our conception of human traits (such as moral thought) themselves [25]. This suggests that compartmentalized reflection on AI governance and neurotechnology governance might not be fully adequate and in fact it might be detrimental if the goal is to identify and fill regulatory gaps.

Conclusions

While few would disagree regarding the importance of addressing the ethical issues raised by neurotechnological and AI innovations, calls for the collaboration of neuroethics and the ethics of AI are sometimes met with some puzzlement. This might be because it is sometimes difficult to determine what such collaboration means. In this paper we have attempted to shed light on such collaboration by focusing on areas that we believe would benefit from joint work: responsible conceptualization, cultural awareness, and governance discussions. We do not rule out the possibility of productive collaborative work in other areas even though more remains to be done in this respect. Importantly, we hope our reflections here will be taken as an opportunity to continue exploring the different ways in which neuroethics and AI ethics can collaborate to ethically shape our future.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Aru J, Larkum ME, Shine JM. The feasibility of artificial consciousness through the lens of neuroscience. Trends Neurosci. 2023;46(12):1008–17. https://doi.org/10.1016/j.tins.2023.09.009.
2. Arun C. AI and the global south: designing for other worlds. In: Dubber M, Pasquale F, Das S, editors. The Oxford handbook of ethics of AI. New York: Oxford Academic; 2020.
3. Balci F, Ben Hamed S, Boraud T, Bouret S, Brochier T, Brun C, Cohen JY, Coutureau E, Deffains M, Doyere V, et al. A response to claims of emergent intelligence and sentience in a dish. Neuron. 2023;111:604–5. https://doi.org/10.1016/j.neuron.2023.02.009.
4. Bassil K. Mending the language barrier: the need for ethics communication in neuroethics. Am J Bioeth Neurosci. 2023;14(4):402–5.
5. Bassil K. The end of 'mini-brains'! Responsible communication of brain organoid research. Mol Psychol. 2024;2:13. https://doi.org/10.12688/molpsychol.17534.2.
6. Berger SE, Rossi F. Addressing neuroethics issues in practice: lessons learnt by tech companies in AI ethics. Neuron. 2022;110(13):2052–6. https://doi.org/10.1016/j.neuron.2022.05.006.
7. Berger SE, Rossi F. AI and neurotechnology: learning from AI ethics to address an expanded ethics landscape. Commun ACM. 2023;66(3):58–68. https://doi.org/10.1145/3529088.
8. Boddington P. Normative modes: codes and standards. In: Dubber M, Pasquale F, Das S, editors. The Oxford handbook of ethics of AI. New York: Oxford Academic; 2020.
9. Borras S, Edler J. The roles of the state in the governance of socio-technical systems transformation. Res Policy. 2020;49:5. https://doi.org/10.1016/j.respol.2020.103971.
10. Bublitz JC. Novel neurorights: from nonsense to substance. Neuroethics. 2022;15(1):7. https://doi.org/10.1007/s12152-022-09481-3.
11. Bublitz JC. What an international declaration on neurotechnologies and human rights could look like: ideas, suggestions. Desiderata AJOB Neurosci. 2023;15:2. https://doi.org/10.1080/21507740.2023.2270512.
12. Butlin P, Long R, Elmoznino E, Bengio Y, Birch J, Constant A, et al. Consciousness in artificial intelligence: insights from the science of consciousness. 2023.
13. Cai H, Ao Z, Tian C, et al. Brain organoid reservoir computing for artificial intelligence. Nat Electron. 2023;6:1032–9. https://doi.org/10.1038/s41928-023-01069-w.
14. Carrozo C. Scientific practice and the moral task of neurophilosophy. AJOB Neurosci. 2019;10(3):115–7.
15. Carrozo C. Conceptual definitions and meaningful generalizability in cognitive enhancement. AJOB Neurosci. 2020;11(4):261–3.
16. Chollet F. ON the measure of intelligence. arXiv:1911.01547v2. 2019. https://doi.org/10.48550/arXiv.1911.01547.
17. Cobb M. The idea of the brain. London: Profile Books; 2021.
18. Coeckelbergh M. AI ethics. Cambridge: The MIT Press; 2020.
19. Colombo M, Piccinini G. The computational theory of mind. Cambridge: Cambridge University Press; 2023. https://doi.org/10.1017/9781009183734
20. Dehaene S, Lau H, Kouider S. What is consciousness, and could machines have it? Science. 2017;358:486–92.
21. Doya K, Ema A, Kitano H, Sakagami M, Russell S. Social impact and governance of AI and neurotechnologies. Neural Networks. 2022;152:542–54. https://doi.org/10.1016/j.neunet.2022.05.012.
22. Evers K, Farisco M, Chatila R, Earp BD, Freire IT, Hamker F, Nemeth E, Verschure PFM, Khamassi M. Artificial consciousness. Some logical and conceptual preliminaries. Philosophy. 2021. https://doi.org/10.48550/arXiv.2403.20177.
23. Evers K, Salles A. Epistemic challenges of digital twins & virtual brains: perspectives from fundamental neuroethics. SCIO J Philos. 2021;21:27–53.
24. Evers K, Salles A, Farisco M. Theoretical framing of neuroethics: the need for a conceptual approach. In: Racine E, Aspler J, editors. Debates about neuroethics: perspectives on its development, focus, and future. Cham: Springer International Publishing; 2017. p. 89–107.
25. Farisco M, Baldassarre G, Cartoni E, Leach A, Petrovici M, Rosemann A, Salles A, Stahl B, van Albada S. A method for the ethical analysis of brain-inspired A.I. Artif Intell Rev. 2024;57:133. https://doi.org/10.1007/s10462-024-10769-4.
26. Farisco M, Evers K, Salles A. On the contribution of neuroethics to the ethics and regulation of artificial intelligence. Neuroethics. 2022;15:4. https://doi.org/10.1007/s12152-022-09484-0.
27. Farisco M, Salles A, Evers K. Neuroethics: a conceptual approach. Camb Q Healthc Ethics. 2018;27(4):717–27. https://doi.org/10.1017/S0963180118000208.
28. Floridi L. The ethics of artificial intelligence: exacerbated problems, renewed problems, unprecedented problems—introduction to the

Special Issue of the American Philosophical Quarterly dedicated to The Ethics of AI 2024. Available at SSRN: https://ssrn.com/abstract=.

29. George D, Lazaro-Gredilla M, Guntupalli JS. From CAPTCHA to commonsense: how brain can teach us about artificial intelligence. Front Comput Neurosci. 2020;14: 554097. https://doi.org/10.3389/fncom.2020.554097.

30. Delegates GNS, et al. Neuroethics questions to guide ethical research in the international brain initiatives. Neuron. 2018;100:1. https://doi.org/10.1016/j.neuron.2018.09.021.

31. Goering S, Yuste R. On the necessity of ethical guidelines for novel neurotechnologies. Cell. 2016;167:3. https://doi.org/10.1016/j.cell.2016.10.029.

32. Goering S, Klein E, SpeckerSullivan L, Wexler A. Recommendations for responsible development and application of neurotechnologies. Neuroethics. 2021;14(3):365–86. https://doi.org/10.1007/s12152-021-09468-6.

33. Hain DS, Jurowetzki R, Squicciarini M, Xu L. Unveiling the neurotechnology landscape: scientific advancements, innovations and major trends. UNESCO. 2023. https://doi.org/10.54678/OCBM4164.

34. Haring S, Mougenot C, Ono F, Watanabe K. Cultural differences in perception and attitude towards robots. Int J Affect Eng. 2014;13(3):149–57.

35. Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-inspired artificial intelligence. Neuron. 2017;95:2. https://doi.org/10.1016/j.neuron.2017.06.011.

36. Herrera Ferra F, Munoz JM, Nicolini H, Zavala GS, Goyri MB. Contextual and cultural perspectives on neurorights: toward and international consensus. AJOB Neurosci. 2023;14(4):3060–368. https://doi.org/10.1080/21507740.2022.2048722.

37. Hopster J, Löhr G. Conceptual engineering and philosophy of technology: amelioration or adaptation? Philos Technol. 2023;36:70. https://doi.org/10.1007/s13347-023-00670-3.

38. Hyun I, Scharf-Deering JC, Lunshof JE. Ethical issues related to brain organoid research. Brain Res. 2020;1732:146653. https://doi.org/10.1016/j.brainres.2020.146653.

39. Ienca M. Neuroethics meets artificial intelligence. In: The Neuroethics Blog. 2019.

40. Ienca M. Democratizing cognitive technology: a proactive approach. Ethics Inf Technol. 2019;21:267–80.

41. Ienca M, Ignatiadis K. Artificial intelligence in clinical neuroscience: methodological and ethical challenges. AJOB Neurosci. 2020;11(2):77–87. https://doi.org/10.1080/21507740.2020.1740352.

42. Illes J, Sahakian BJ. Preface. In: Illes J, Sahakian BJ, editors. The Oxford handbook of neuroethics. Oxford: Oxford University Press; 2011.

43. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1:389–99. https://doi.org/10.1038/s42256-019-0088-2.

44. Johnson D, Verdicchio M. Reframing AI discourse. Mind Mach. 2017;27:575–90. https://doi.org/10.1007/s11023-017-9417-6.

45. Johnson LSM, Rommelfanger KS. The Routledge handbook of neuroethics. New York: Routledge, Taylor & Francis Group; 2018.

46. Johnson W. Catching up with convergence: strategies for bringing together the fragmented regulatory governance of brain–machine interfaces in the United States. Ann Health Law Life Sci. 2021;30:1.

47. Kagan BJ, Razi A, Bhat A, Kitchen AC, Tran NT, Habibollahi F, Khajehnejad M, Parker BJ, Rollo B, Friston KJ. Scientific communication and the semantics of sentience. Neuron. 2023;111:606–7. https://doi.org/10.1016/j.neuron.2023.02.008.

48. Kataoka M, Gyngell C, Savulescu J, Sawai T. Chapter eleven—moral dimensions of synthetic biological intelligence: unravelling the ethics of neural integration. In: Ienca M, Starke G, editors. Developments in neuroethics and bioethics, vol. 7. New York: Academic Press; 2024. p. 205–19.

49. Kellmeyer P. Artificial intelligence in basic and clinical neuroscience: opportunities and ethical challenges. Neuroforum. 2019;25(4):241–50. https://doi.org/10.1515/nf-2019-0018.

50. Lefley HP. Families, culture, and mental illness: constructing new realities. Psychiatry Interpersonal Biol Process. 1998;61(4):335–55. https://doi.org/10.1080/00332747.1998.11024846.

51. Levy N. Neuroethics. Cambridge: Cambridge University Press; 2007.

52. Levy N. The value of consciousness. J Conscious Stud. 2014;21(1–2):127–38.

53. Lupton D. Language matters: the 'digital twin' metaphor in health and medicine. J Med Ethics. 2021;47:409. https://doi.org/10.1136/medethics-2021-107517.

54. Llorca Albareda J, Rueda J. Divide and rule? Why ethical proliferation is not so wrong for technology ethics. Philos Technol. 2023;36:10. https://doi.org/10.1007/s13347-023-00609-8.

55. Makari G. Soul machine: the invention of the modern mind. New York: W.W.Norton and Co; 2017.

56. Marchant G, Tournas L. Filling the governance gap: international principles for responsible development of neurotechnologies. AJOB Neurosci. 2019;10(4):176–8. https://doi.org/10.1080/21507740.2019.1665135.

57. Marchant G. The growing gap between emerging technologies and the law. In: Marchant G, Allenby B, Herkert J, editors. The growing gap between emerging technologies and legal-ethical oversight the international library of ethics, law and technology. Dordrecht: Springer; 2011. p. 19–33. https://doi.org/10.1007/978-94-007-1356-7_2.

58. Marchant G. Governance of emerging technologies as a wicked problem. Vanderbilt Law Rev. 2020;73(6):1861–77.

59. Marchiori S, Scharp K. What is conceptual disruption? Ethics Inf Technol. 2024;26:18. https://doi.org/10.1007/s10676-024-09749-7.

60. Marcus S, Charles A. Dana foundation. Neuroethics: mapping the field: conference proceedings, May 13–14, 2002, San Francisco, California, New York: Dana Press.

61. Northoff G. What is neuroethics? Empirical and theoretical neuroethics. Curr Opin Psychiatry. 2009;6:565–9. https://doi.org/10.1097/YCO.0b013e32832e088b.

62. O'Shaughnessy MR, Johnson WG, Tournas LN, Rozell CJ, Rommelfanger KS. Neuroethics guidance documents: principles, analysis, and implementation strategies. J Law Biosci. 2023;10(2):lsad25. https://doi.org/10.1093/jlb/lsad025.

63. OECD. Recommendation of the council on responsible innovation in neurotechnology. In: OECD/LEGAL/0457. 2019.

64. OhEigeartaigh SS, Whittlestone J, Liu Y, Zeng Y, Liu Z. Overcoming barriers to cross-cultural cooperation in AI ethics and governance. Philos Technol. 2020;33:571–93. https://doi.org/10.1007/s13347-020-00402-x.

65. Pereira A Jr. The role of sentience in the theory of consciousness and medical practice. J Conscious Stud. 2021;28(7–8):22–50.

66. Pichl A, Ranisch R, Altinok OA, Antonakaki M, Barnhart AJ, Bassil K, Boyd JL, Chinaia AA, Diner S, Gaillard M, Greely HT, Jowitt J, Kreitmair K, Lawrence D, Lee TN, McKeown A, Sachdev V, Schicktanz S, Sugarman J, Trettenbach K, Wiese L, Wolff H, Árnason G. Ethical, legal and social aspects of human cerebral organoids and their governance in Germany, the United Kingdom and the United States. Front Cell Dev Biol. 2023. https://doi.org/10.3389/fcell.2023.1194706.

67. Powers TM, Ganascia JG. The ethics of the ethics of AI. In: Dubber M, Pasquale F, Das S, editors. The Oxford handbook of ethics of AI. New York: Oxford Academic; 2020.

68. Racine E, Senghor AS. Diversity in neuroethics: which diversity and why it matters? In: Farisco M, editor. Neuroethics and cultural diversity. London: ISTE-Wiley; 2023. p. 64–5.

69. Rainey S, Erden Y. Correcting the brain? The convergence of neuroscience, neurotechnology, psychiatry, and artificial intelligence. Sci Eng Ethics. 2020;2020(26):2439–54. https://doi.org/10.1007/s11948-020-00240-2.

70. Reber AS. The first minds: caterpillars, karyotes, and consciousness. New York: Oxford University Press; 2019.

71. Resseguier A, Rodrigues R. Ethics as attention to context: recommendations for the ethics of artificial intelligence [version 2; peer review: 1 approved, 2 approved with reservations]. Open Res Europe. 2021;1:27. https://doi.org/10.12688/openreseurope.13260.2.

72. Rizk N. Artificial intelligence and inequality in the Middle East: the political economy of inclusion. In: Dubber M, Pasquale F, Das S, editors. The Oxford handbook of ethics of AI. New York: Oxford Academic; 2020.

73. Robinson JT, Rommelfanger KS, Anikeeva PO, Etienne A, French J, Gelinas J, Grover P, Picard R. Building a culture of responsible neurotech: neuroethics as socio-technical challenges. Neuron. 2022;110(13):2057–62. https://doi.org/10.1016/j.neuron.2022.05.005.

74. Rommelfanger KS, Ramos K, Salles A. Conceptual conundrums for neuroscience. Neuron. 2023;111(5):608–9. https://doi.org/10.1016/j.neuron.2023.02.016.

75. Rommelfanger K, Specker-Sullivan L. The dilemma of cross-cultural neuroethics. In: Farisco M, editor. Neuroethics and Cultural Diversity. London: Wiley; 2023.

76. Sætra HS, Danaher J. To each technology its own ethics: the problem of ethical proliferation. Philos Technol. 2022;35:93. https://doi.org/10.1007/s13347-022-00591-7.

77. Salles A. Neuroethics and culture. In: Farisco M, editor. Neuroethics and cultural diversity. London: ISTE Ltd; 2023.

78. Salles A. Some reflections on the neurorights debate. In: Navarro MS, Dura-Bernal S, Gulotta CM, editors. The risks and challenges of neurotechnologies for human rights. UNESCO, University of Milan-Bicocca, SUNY Downstate; 2023.

79. Salles A, Evers K, Farisco M. The need for a conceptual expansion of neuroethics. AJOB Neurosci. 2019;10(3):126–8. https://doi.org/10.1080/21507740.2019.1632972.

80. Salles A, Evers K, Farisco M. Anthropomorphism in AI. AJOB Neurosci. 2020;11(2):88–95. https://doi.org/10.1080/21507740.2020.1740350.

81. Sawai T, Hayashi Y, Niikawa T, Shepherd J, Thomas E, Lee TL, Erler A, Watanabe M, Sakaguchi H. Mapping the ethical issues of brain organoid research and application. AJOB Neurosci. 2022;13(2):81–94. https://doi.org/10.1080/21507740.2021.1896603.

82. Sayed A. Conceptualization of mental illness within Arab cultures: meeting challenges in cross-cultural settings. Social Behav Personal Int J. 2003;31(4):333–41.

83. Schermer M. The mind and the machine. On the conceptual and moral implications of brain–machine interaction. NanoEthics. 2009;3(3):217–30. https://doi.org/10.1007/s11569-009-0076-9.

84. Schermer M. The Cyborg fear: how conceptual dualisms shape our self-understanding. AJOB Neurosci. 2014;5(4):56–7. https://doi.org/10.1080/21507740.2014.951784.

85. Schwitzgebel E. AI systems must not confuse users about their sentience or moral status. Patterns (N Y). 2023;4(8):100818. https://doi.org/10.1016/j.patter.2023.100818.

86. Searle JR. Biological naturalism. In: Velmans M, Schneider S, editors. The Blackwell companion to consciousness. Malden: Blackwell Publishing Ltd; 2007. p. 325–34.

87. Shook JR, Giordano J. A principled and cosmopolitan neuroethics: considerations for international relevance. Philos Ethics Humanit Med. 2014;9:1. https://doi.org/10.1186/1747-5341-9-1.

88. Smirnova L, Caffo BS, Gracias DH, Huang Q, Morales Pantoja IE, Tang B, Zack DJ, Berlinicke CA, Boyd JL, Harris TD, Johnson EC, Kagan BJ, Kahn J, Muotri AR, Paulhamus BL, Schwamborn JC, Plotkin J, Szalay AS, Vogelstein JT, Worley PF, Hartung T. Organoid intelligence (OI): the new frontier in biocomputing and intelligence-in-a-dish. Front Sci. 2023. https://doi.org/10.3389/fsci.2023.1017235.

89. Stix C. Actionable principles for artificial intelligence policy: three pathways. Sci Eng Ethics. 2021;27:15. https://doi.org/10.1007/s11948-020-00277-3.

90. Suzuki M, Pennartz CMA, Aru J. How deep is the brain? The shallow brain hypothesis. Nat Rev Neurosci. 2023;24(12):778–91. https://doi.org/10.1038/s41583-023-00756-z.

91. Ullman S. Using neuroscience to develop artificial intelligence. Science. 2019;363(6428):692–3.

92. Ulnicane I, Leach T, Knight W, Stahl BC, Wanjiku WG. Framing governance for a contested emerging technology: insights from AI policy. Policy and Society. 2021;40(2):158–77. https://doi.org/10.1080/14494035.2020.1855800.

93. Voigtlaender S, Pawelczyk J, Geiger M, Vaios EJ, Karschnia P, Cudkowicz M, Dietrich J, Hebold Haraldsen IRJ, Feigin V, Owolabi M, White TL, Świeboda P, Farahany N, Natarajan V, Winter SF. Artificial intelligence in neurology: opportunities, challenges, and policy implications. J Neurol. 2024;271:2258–73. https://doi.org/10.1007/s00415-024-12220-8.

94. Wallach W, Marchant G. Toward the agile and comprehensive International Governance of AI and robotics. Proc IEEE. 2019;107(3):505–8.

95. Wexler A, Reiner PB. Oversight of direct-to-consumer neurotechnologies. Science. 2019;363:234–5. https://doi.org/10.1126/science.aav0223.

96. Yuste R, Genser J, Herrmann S. It's time for neuro-rights: new human rights for the age of neurotechnology. Horizons. 2021:154–164.

Publisher's Note