1 Introduction: Ethical Issues of Neurotechnologies and Artificial Intelligence

Orsolya Friedrich and Andreas Wolkenstein

Contents

Abstract

In this introduction to the volume, we present an overview of existing research on intelligent neurotechnologies, i.e., the combination of neurotechnologies with Artificial Intelligence (AI). Further, we present the ideas behind this volume and an overview of each chapter.

1.1 Neurotechnology + Artificial Intelligence = Intelligent Neurotechnologies (INT)

Imagine that the coffee machine in your kitchen starts brewing your urgently needed morning coffee as soon as you *think* the command "start the coffee machine" while you are still in bed. Is that realistic? Is it desirable? Using neurotechnologies, i.e.,

O. Friedrich (✉)
Institute of Philosophy, FernUniversität in Hagen, Hagen, Germany
e-mail: orsolya.friedrich@fernuni-hagen.de

A. Wolkenstein
Institute of Ethics, History and Theory of Medicine, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany
e-mail: andreas.wolkenstein@med.uni-muenchen.de

O. Friedrich et al. (eds.), *Clinical Neurotechnology meets Artificial Intelligence*, Advances in Neuroethics, https://doi.org/10.1007/978-3-030-64590-8_1

technologies that lead to understanding, changing or interacting with the brain, combined with artificial intelligence (AI) might allow for such an application, even though many scientists doubt that technologies such as this one could be available in the near future. However, basic principles of brain-computer interfacing (BCI) have become reality and are currently the subject of intense research efforts [1–4]. BCIs measure brain activity and convert brain signals into computer commands, e.g., moving a cursor or a wheelchair [5, 6]. The most common way to measure brain activity is with non-invasive electroencephalography (EEG). BCIs use the power of thought or of focusing on a signal in order to give computational commands and require no neuromuscular innervation.

At the same time, BCIs and other neurotechnologies stand in relation with another emerging technology: AI. AI is already being used in many technologies to solve problems, which usually require human intelligence, such as reasoning, planning, and speech perception [7]. It is not a technology designed for a specific task, but cuts across all societal domains [8, 9] and comprises several technologies such as machine learning and artificial neural networks. The term "AI" thus denotes a variety of converging technologies that are used across many platforms and technologies. Kellmeyer [10] lists five different aspects: ubiquitous data collection, storage and processing of large amounts of data (big data), high performance analysis, machine learning, and interfaces for human-AI interaction.

AI is used in a number of ways in neuroscience and neurotechnology in the medical domain [11]. For example, computer vision capacities are being applied to detect tumors in magnetic resonance imaging (MRI) [12] or to detect anomalies in other kinds of data [13], e.g., EEG data [14–16]. These capacities lead to an improved diagnosis, prediction, and treatment of clinical pictures in a variety of medical domains [10]. In psychiatry, researchers have recently used AI to reach a biomarker-based diagnosis and determine therapy in patients with dementia, attention deficit hyperactivity disorder (ADHD), schizophrenia, autism, depression, and posttraumatic stress disorder (PTSD) [17–20]. AI that is used for speech recognition, in addition to many available data sources on the internet, helps researchers predict mental illness, for example [21].

Beyond its application in clinical research and therapy, AI is being used in combination with neurotechnologies. Big data and deep learning, for example, are promising trends that will influence the development of BCIs [22]. Among many other uses, these devices can be used by patients who suffer from amyotrophic lateral sclerosis (ALS) or severe paralysis in order to restore communication capacities and mobility, or in rehabilitation to facilitate the recovery process of patients after stroke [23–25]. With the help of AI, important BCI features such as signal processing and feature extraction can be improved [22]. Outside the strictly medical arena, EEG-based BCIs and other forms of AI-based neurotechnology are sold for entertainment purposes [26]. Facebook famously works with a typing-by-brain technology, which allows for a seamless social media experience [27]. Research behind this technology was already capable of showing how algorithms could decode speech in real time with a high amount of reliability [28]. Similarly, progress has been made in terms of facial recognition in EEG data [29]. BCIs, as well as other applications of (AI-enhanced) neurotechnology can also be found in military research. Warfighter enhancement is one motivation, but others include enhancing military equipment or deception detection [30–33].

In addition to technological development and progress, the number of articles, books, and events such as workshops or conferences that deal with the neuroethics of AI and neurotechnology is steadily increasing. Generally speaking, AI raises a host of original problems that can most aptly be summarized as "black box"-problems: It becomes increasingly difficult to supervise and control an AI's operation, because it manages its decision-making logic all by itself [34–37]. The combination of neurotechnologies and AI raises a host of further pressing problems. Yuste and colleagues [38] mention four broad areas of ethical concern: privacy and consent, agency and identity, augmentation, as well as bias. They propose various measures to address these issues, ranging from technological safeguards to legislation. For medical neurotechnology, a number of articles also emphasized problems regarding data protection and privacy as important issues to consider [39]. Moreover, questions of responsibility and shared agency are repeatedly brought up when it comes to neurotechnologies [40]. How BCIs affect agency and autonomy is another topic that drew attention to philosophers and ethicists [41, 42]. This body of research adds to more general approaches that examine the ethical quality of algorithms per se [9, 43]. Articles on issues such as hackability and problems derived from unwanted access to brain data [44] complement work that looks at specific forms of neurotechnology, e.g., in the medical, military, or consumer area [32, 33, 45, 46]. In addition, neurotechnology becomes increasingly interesting for political philosophers and others who approach INT with an eye on regulation questions and broader democratic worries [39, 47].

1.2 Novel Philosophical, Ethical, Legal, and Sociological Approaches to INT: An Overview

As this brief overview shows, many questions have already been addressed in the emerging literature both on technical issues and the normative implications of INT. Some of these questions have not been sufficiently or satisfyingly answered. Scholars from philosophy, sociology, and the law continue to exchange arguments and ideas while medical researchers, engineers, and computer scientists keep exploring new technologies and improve existing ones. The aim of this book is to provide a forum for the continuous exchange of these arguments and ideas. From a philosophical and ethical perspective, normatively relevant notions such as agency, autonomy, or responsibility have to be analyzed if humans interact with INT. This volume also asks, in a descriptive manner, how the reality of using INT would look like. It sheds light on the legal dimensions of INT. In addition, it explores a number of specific use cases, in that these concrete scenarios reveal more about the various domains of human agency in situations where technology and human-machine interaction play a distinctive role.

Accordingly, the methods used in this book vary considerably. They range from philosophical analysis, sociologically inspired descriptions, legal analysis, and socio-empirical research. This provides the book with the capacity to address a wide range of philosophical, normative, social, legal, and empirical dimensions of neurotechnology and AI. Most of the papers of this volume are the result of a conference that was held in Munich, in which the ethics of (clinical) neurotechnologies and AI were intensely discussed.[1]

The *first section* of the book reflects on some philosophically relevant phenomena and implications of neurotechnology use. From a philosophical and ethical perspective, it must be asked how normatively relevant notions such as action, agency, autonomy, or responsibility can be conceptualized if humans act and interact with neurotechnologies. The most basic question is if BCI effects are actions at all and if there are normatively relevant differences between paradigmatic bodily actions and BCI-mediated actions. If there is no action or agency to be claimed, subsequent issues of autonomy and responsibility are affected, as well. Therefore, philosophical analyses of BCI use that focus on action-theoretical implications have emerged recently [41]. Two articles in this first section take this path.

Tom Buller analyzes the implications of BCI use for the nature of action. He claims that present BCI-mediated behavior fails to meet the necessary condition of intentional actions, namely the causation of an event and thus of bodily movement that is directly related to relevant beliefs and desires. Furthermore, he states that current BCI-mediated changes in the world do not qualify as non-deviant causal processes.

Sebastian Drosselmeier and *Stephan Sellmaier* also address the issue of action. However, they focus on the acquisition of a skill while using BCIs, which allows the user to make BCI-mediated changes in the world without performing a mental act. This would result—according to their argumentation—in the ability to perform BCI actions as basic actions. They also conclude that BCI users are able to differentiate between having a thought and an action relevant intention. Therefore, skilled users should be seen as competent and able to voluntarily control the BCI effects, which they cause in the world.

The concepts of action and agency are closely connected to the concept of autonomy. Therefore, this suggests that some authors have recently also addressed the implications of BCI use on autonomy [42]. The first section of this volume also deals with this issue. Realizing the ability to act autonomously might be hampered or enhanced by using neurotechnologies.

Anna Wilks takes a closer look at the question of whether it would be a paradox or a possibility, following Kant, to augment autonomy through neurotechnologies. The paradox seems obvious at first hand: someone claims to augment autonomy with BCI use, but is able to perform self-legislation, whereas autonomous agency in a Kantian understanding requires that the person is not affected by external factors. Wilks, however, suggests that operating with a broader Kantian framework would

[1] https://neurotechmeetsai.wordpress.com/

allow integrating external components of BCIs into the understanding of self-legislation and thus avoid the paradox.

Pim Haselager, *Giulio Mecacci,* and *Andreas Wolkenstein* argue that BCIs, especially passive BCIs, shed new light on the traditional question of agency in philosophy. More precisely, they argue that the notion of ownership of action ("was that me?") might be affected by closely examining the action-theoretical implications of passive BCIs. If BCIs register intentions without the user being aware of this, and if they consequently act on them, then subconscious brain states may influence one's actions in a technology-mediated way. This observation serves as the basis for their plea to use passive BCIs, or what they call symbiotic technology, in experimentally guided thought experiments aimed at the study of the notion of agency. The authors suspect that by doing so, symbiotic technology may give new answers to how we must understand ownership of action and what consequences we have to expect.

Andreas Wolkenstein and *Orsolya Friedrich* contribute to the first section of the volume by summarizing the philosophical and ethical analysis that they described in their BCI-use analyzing project (Interfaces) and suggest some future directions for research and regulation of BCI development and use. They show that relevant results have been produced in recent philosophical, ethical, social, and legal reflections of BCI use. However, concluding results that could profoundly advise technology-regulating institutions or engineers are not present yet. Nevertheless, the development of AI-driven neurotechnologies are emerging and therefore, some preliminary ethically based regulatory framework is necessary. They suggest using procedural criteria as a first step.

Neurotechnology and AI also have broad social implications. These social implications not only include societal issues in general; certain areas of society, like research and medicine, are affected in a specific way. The *second section* of this volume focuses on some social implications of neurotechnology and AI use.

Matthew Sample and *Eric Racine* recall in their article that other emerging technologies, e.g., genomics or nanotechnology, have been promoted in ethics research in the past similar to the way that neural technologies are now. They address the question of how ethics researchers should deal with such research developments and question the significance of digital society for ethics research. They show how the significance of artificial intelligence and neural technologies, as examples of digital technologies, is affected by both sociological and ethical factors. They conclude that ethics researchers have to be careful in attributing significance and to reflect their own function in the process of attribution.

Johannes Kögel also focuses on BCI use from a sociological perspective. He shows that the BCI laboratory is not only a place to train this novel technology, but also a place of crisis management. The aim to discuss BCI use also as crisis management is to understand this social process and to increase sensitivity for the user experience. He argues that users currently experience BCI training and tasks as tedious and exhausting, because they have to make many "back-to-back decisions" for a long period of time and under immense time pressure, which is not common to activities in everyday life. His focus emphasizes the importance of developing BCI applications that allow for a more routine way of acting.

Jennifer R. Schmid and *Ralf J. Jox* further highlight the relevance and implications of the training process for the user experience in BCIs. They report on a qualitative interview study with healthy BCI users, e.g., neuro-gamers or pilots. The interviews show that the success of BCI use strongly depends on the motivation as well as the duration of training and that the time-consuming procedure of use results in discomfort and cognitive exhaustion.

This *second section* of this volume also approaches intelligent neurotechnologies from a legal perspective. The legal system faces the need to update some of its notions and regulatory action is needed to cover these new, neurotechnology-based forms of acting and acting together. BCIs also raise the question about mental privacy as well as data and consent issues.

Susanne Beck focuses on criminal law issues that result from neurotechnology use. She shows how neurotechnologies might lead to diffusion on the end of the victim, as well as the offender. Such diffusion would be important for criminal law, in that in traditional criminal law the roles of offender and victim are very clear. Therefore, criminalizing might lose some of its legitimacy. Another problematic diffusion in criminal law might occur, if there are no clear borders between the body and the mind.

Stephen Rainey et al. address further legally relevant issues, namely those related to data and consent in neural recording devices. They discuss whether current data protection regulation is adequate. They conclude that brain-reading devices present difficult consent issues for consumers and developers of the technology. They are also a potential challenge for current European data protection standards. Their use might become legally problematic, if the nature of the device results in an inability for the user to exercise their rights.

Finally, in the *third section* the book takes a closer look at neurotechnologies in their contexts of use. This section covers both the introduction of using neurotechnologies in various domains and an explication and discussion of their deeper philosophical, ethical, and social implications.

Ralf J. Jox discusses the ethical implications of the use of neurotechnologies and AI in the domain of medicine. He shows that such technology use challenges not only the patient–physician relationship, but also the whole character of medicine. He further highlights the potential threats to human nature, human identity, and the fundamental distinction between human beings and technological artifacts that could arise when AI technology with certain features is closely connected with the human brain.

The next contribution highlights one of these close connections of AI-neurotechnology and the human brain. *Stephen Rainey* discusses neuro-controlled speech neuroprosthesis from an ethical perspective. A speech neuroprosthesis picks out linguistically relevant neural signals in order to synthesize and realize, artificially, the overt speech sounds that the signals represent. The most important question in this special neurotechnology application is whether the synthesized speech represents the user's speech intentions and to what extent he can control the speech neuroprosthesis.

Georg Starke's contribution addresses another field of clinical neuroscience, namely the application of ML to neuroimaging data and the potential challenges of this application with regard to transparency and trust. He shows why transparency and trustworthiness are not necessarily linked and why transparency alone won't solve all the challenges of clinical ML applications.

Another field of application of neurotechnology and AI is their use in the military. *Jean-Marc Rickli* and *Marcello Ienca* discuss the security and military implications of neurotechnology and AI with regard to five security-relevant issues, namely data bias, accountability, manipulation, social control, weaponization, and democratization of access. They show that neurotechnology and AI both raise security concerns and share some characteristics: they proliferate outside supervised research settings, they are used for military aims, and they have a transformative and disruptive character. They highlight that it is extremely difficult to control the use and misuse of these technologies and call for global governance responses that are able to deal with the special characteristics of these technologies.

Finally, *Mathias Vukelić* directs our attention to a new research agenda for designing technology. Given the increasingly symbiotic nature of neurotechnology, where humans and technology closely interact, he emphasizes the need for a human-centered approach that puts human needs at the core. He attests that the detection of brain states, such as emotional or affective reactions, are of great potential for the development of symbiotic, interactive machines. Beyond assistive technology, this research leads to neuroadaptive technologies that are usable in a broad variety of domains. Vukelić argues that the primary goal of such an undertaking is the alignment of increasingly intelligent technology with human needs and abilities. While this could itself be viewed as following an ethical imperative, the author also stresses the wider ethical and societal implications of such a research agenda.

This short overview of existing research on intelligent neurotechnologies and of the articles in this volume offers a first insight into the emerging philosophical, ethical, legal, and social difficulties that we will have to face in the future and which require further conceptual as well as empirical research.

Acknowledgments Work on this paper was funded by the Federal Ministry of Education and Research (BMBF) in Germany (INTERFACES, 01GP1622A) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—418201802. We would like to thank Dorothea Wagner von Hoff for *proof reading* the article, Meliz-Sema Kaygusuz and Bernadette Scherer for formatting.

References

1. McFarland DJ, Wolpaw JR. EEG-based brain-computer interfaces. Curr Opin Biomed Eng. 2017;4:194–200.
2. Clerc M, Bougrain L, Lotte F, editors. Brain-computer interfaces 1. Foundations and methods. London: Wiley; 2016.
3. Clerc M, Bougrain L, Lotte F, editors. Brain-computer interfaces 2. Technology and applications. London: Wiley; 2016.

4. Graimann B, Allison B, Pfurtscheller G, editors. Brain-computer interfaces. Revolutionizing human-computer interaction. Berlin: Springer; 2010.
5. Shih JJ, Krusienski DJ, Wolpaw JR. Brain-computer interfaces in medicine. Mayo Clin Proc. 2012;87(3):268–79.
6. Graimann B, Allison B, Pfurtscheller G. Brain-computer interfaces: a gentle introduction. In: Graimann B, Allison B, Pfurtscheller G, editors. Brain-computer interfaces: revolutionizing human-computer interaction. Berlin: Springer; 2010. p. 1–27.
7. Luxton DD, editor. Artificial intelligence in behavioral and mental health care. Amsterdam: Elsevier; 2016.
8. Mainzer K. Künstliche Intelligenz. Wann übernehmen die Maschinen? Berlin: Springer; 2016.
9. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data Soc. 2016;3(2):1–21.
10. Kellmeyer P. Artificial intelligence in basic and clinical neuroscience: opportunities and ethical challenges. e-Neuroforum. 2019;25(4):241–50.
11. Gunes O, Gunes G, Seyitoglu DC. The use of artificial intelligence in different medical branches: an overview of the literature. Med Sci. 2019;8(3):770–3.
12. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging. 2016;35(5):1240–51.
13. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep. 2016;6(1):26286.
14. Tabar YR, Halici U. A novel deep learning approach for classification of EEG motor imagery signals. J Neural Eng. 2016;14(1):016003.
15. Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum Brain Mapp. 2017;38(11):5391–420.
16. Schirrmeister RT, Gemein L, Eggensberger K, Hutter F, Ball T. P64. Deep learning for EEG diagnostics. Clin Neurophysiol. 2018;129(8):e94.
17. Fakhoury M. Artificial intelligence in psychiatry. In: Kim Y-K, editor. Frontiers in psychiatry: artificial intelligence, precision medicine, and other paradigm shifts. Singapore: Springer Singapore; 2019. p. 119–25.
18. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci Biobehav Rev. 2017;74:58–75.
19. Meyer-Lindenberg A. Künstliche Intelligenz in der Psychiatrie – ein Überblick. Nervenarzt. 2018;89(8):861–8.
20. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
21. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci. 2018;115(44):11203.
22. Lin C, Liu Y, Wu S, Cao Z, Wang Y, Huang C, et al. EEG-based brain-computer interfaces: a novel neurotechnology and computational intelligence method. IEEE Syst Man Cybernet Mag. 2017;3(4):16–26.
23. Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain–computer interfaces for communication and rehabilitation. Nat Rev Neurol. 2016;12(9):513–25.
24. Salisbury DB, Parsons TD, Monden KR, Trost Z, Driver SJ. Brain–computer interface for individuals after spinal cord injury. Rehabil Psychol. 2016;61(4):435–41.
25. McFarland DJ, Daly J, Boulay C, Parvaz MA. Therapeutic applications of BCI technologies. Brain Comput Interfaces. 2017;4(1–2):37–52.
26. Blankertz B, Acqualanga L, Dähne S, Haufe S, Schultze-Kraft M, Sturm I, et al. The Berlin brain-computer interface: progress beyond communication and control. Front Neurosci. 2016;10:1–24.

27. Robertson A. Facebook just published an update on its futuristic brain-typing project. The Verge 2019 July 30.
28. Moses DA, Leonard MK, Makin JG, Chang EF. Real-time decoding of question-and-answer speech dialogue using human cortical activity. Nat Commun. 2019;10(1):3096.
29. Nemrodov D, Niemeier M, Patel A, Nestor A. The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction. eNeuro. 2018;5(1):ENEURO.0358-17.2018.
30. Evans NG, Moreno JD. Neuroethics and policy at the National Security Interface: a test case for neuroethics theory and methodology. In: Racine E, Aspler J, editors. Debates about neuroethics. Cham: Springer; 2017. p. 141–57.
31. Tennison MN, Moreno JD. Neuroscience, ethics, and national security: the state of the art. PLoS Biol. 2012;10(3):e1001289.
32. Kotchetkov IS, Hwang BY, Appelboom G, Kellner CP, Connolly ES. Brain-computer interfaces: military, neurosurgical, and ethical perspective. Neurosurg Focus. 2010;28(5):E25.
33. Munyon CN. Neuroethics of non-primary brain computer interface: focus on potential military applications. Front Neurosci. 2018;12:696.
34. de Laat PB. Algorithmic decision-making based on machine learning from big data: can transparency restore accountability? Philos Technol. 2018;31(4):525–41.
35. Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. Sci Robot. 2017;2:eaan6080.
36. Pasquale F. The black box society. The secret algorithms that control money and information. Cambridge: Harvard University Press; 2015.
37. O'Neil C. Weapons of math destruction. How big data increases inequality and threatens democracy. New York: Crown; 2016.
38. Yuste R, Goering S, Agüera y Arcas B, Bi G, Carmena JM, Carter A, et al. Four ethical priorities for neurotechnologies and AI. Nature. 2017;551(7679):159–63.
39. Kellmeyer P. Big brain data: on the responsible use of brain data from clinical and consumer-directed neurotechnological devices. Neuroethics. 2018. https://doi.org/10.1007/s12152-018-9371-x.
40. Bublitz C, Wolkenstein A, Jox RJ, Friedrich O. Legal liabilities of BCI-users: responsibility gaps at the intersection of mind and machine? Int J Law Psychiatry. 2019;65:101399.
41. Steinert S, Bublitz C, Jox R, Friedrich O. Doing things with thoughts: brain-computer interfaces and disembodied agency. Philos Technol. 2019;32(3):457–82.
42. Friedrich O, Racine E, Steinert S, Pömsl J, Jox RJ. An analysis of the impact of brain-computer interfaces on autonomy. Neuroethics. 2018. https://doi.org/10.1007/s12152-018-9364-9.
43. Wolkenstein A, Jox RJ, Friedrich O. Brain–computer interfaces: lessons to be learned from the ethics of algorithms. Camb Q Healthc Ethics. 2018;27(4):635–46.
44. Ienca M, Haselager P. Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. Ethics Inf Technol. 2016;18(2):117–29.
45. Ienca M, Haselager P, Emanuel EJ. Brain leaks and consumer neurotechnology. Nat Biotechnol. 2018;36:805.
46. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res. 2019;21(5):e13216.
47. Wolkenstein A, Friedrich O. Brain-computer interfaces: current and future investigations in the philosophy and politics of neurotechnology. In: Friedrich O, Wolkenstein A, Bublitz C, Jox RJ, Racine E, editors. Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications. Heidelberg: Springer; 2021.