Chapter 11
Brain-Machine Interfaces and Artificial Intelligence: An Ethical Overview

Steven S. Gouveia[1]

This chapter critically examines the ethical implications of integrating Artificial Intelligence (AI) with Brain-Machine Interfaces (BMIs), focusing on specific concerns related to safety, autonomy and responsibility, and technical aspects of these technologies. The development of BMIs in connection to AI algorithms enable more accurate decoding of neural signals, facilitating applications in neuroprosthetics, cognitive enhancement, and even direct brain-to-brain communication. Even though this is positive and can help numerous patients with specific diseases, several raises ethical concerns should be raised: (i) BMIs can pose a direct risk of harm, particularly for invasive methods that require surgical interventions related with implants, since there is potential for complications including infections surrounding the tissue or acute trauma to the brain (Wolpaw and Wolpaw, 2012); for long-term implants, the affected neural tissue may also develop glial scarring, which can surround the implant and obstruct the BMI function (Borton et al., 2013); (ii) issue regarding autonomy and responsibility could also arise from BMIs: is an action primarily or exclusively generated by a

[1] Research Fellow, Mind Language and Action Group, Institute of Philosophy, University of Porto, Portugal. Honorary Professor, Faculty of Medicine, University Andrés Bello, Chile.

device still attributable to a human? If a BMI device significantly influences an individual's decision-making process, this could potentially undermine autonomy; for example, studies found that the illusion of agency is a reality, where BMI patients inaccurately claim to be the agent of action when that was not the case (Vlek et al., 2014); finally, (iii) there are several technical issues that needs to be considered from a normative point of view: BMIs rely on accurately capturing neural signals: noise, signal degradation, and interference can affect the quality and reliability of the signals, leading to inaccurate or unreliable control of devices; over time, neural signals can change due to various factors, such as electrode drift or biological changes, and this can lead to a decrease in the accuracy and stability of the BMI's performance; moreover, interference between neighbouring electrodes can lead to crosstalk, where the signals from one electrode affect the measurements from others. Neural crosstalk can complicate accurate signal interpretation (Clausen, 2011). The chapter concludes by addressing the need for transparent AI algorithms in BMIs, advocating for ethical oversight to ensure that these innovations prioritize sufficient ethical standards, while balancing the benefits and risks of neurotechnology.

Keywords: Brain-Machine Interface, Artificial Intelligence, Ethics, Autonomy, Privacy.

1. Introduction

1.1 Background context

The fusion of Brain-Machine Interfaces (BMIs) and Artificial Intelligence (AI) marks, for some, a key moment in the development of neurotechnology broadly conceived: these systems are no longer mere lab experiments, but are here and now entering clinical trials and towards human-based applications with an interesting cognitive and therapeutical potential (van Stuijvenberg et al., 2024a). Fan et al. (2023) reported stable decoding accuracy of 93.8% over 403 days – more than one year – in a real-time BCI system (intracortical handwriting), making use of continual online recalibration framework, which showed a relevant performance in a human participant, underscoring mutually their technical feasibility and clinical capability.

Nevertheless, this technical progress carries ethical complexities: the potential to be able to decode human intentions or emotions with these kinds of interfaces make us confronted with several philosophical questions about personal identity, privacy, and autonomy: for example, Yuste and Goering (2017) argue that these technologies risk undermining fundamental human values, demanding frameworks that can protect privacy, identity, agency, and equality. In the same line of thought, van Stuijvenberg et al. (2024b) underline that user control ought to be highly relevant from an ethical point of view: users must not be instrumentalized, but their agency ought to be explicit, controllable and clear for both themselves and other people.

Moreover, national-level ethical efforts have been rising to lead what can be called “responsible” BMI deployment: for example, China’s 2024 Guidelines for Research Ethics in Brain-Computer Interface (NRECChina, 2024) advocate sustainable governance via alignment with already established medical protocols (e.g., the Declaration of Helsinki), while UNESCO and European bodies are concurrently developing similar frameworks (UNESCO, 2024). These initiatives indicate a growing concern focused on the idea that AI-inspired BMI technologies in general cannot overtake the ethical and legal frameworks meant to steward it. The aim of this chapter will be to explore how, in different contexts, AI-driven BMIs may be already overtaking some fundamental ethical concerns.

1.2. Definitions and Scope

Before starting the ethical analysis, let's make it clear what kind of technologies this chapter will be focusing on. First, Brain–Machine Interfaces (BMIs, also referred in the literature as BCIs) are systems that can record activity from the central nervous system (whether via implanted, invasive sensors, or non-invasive scalp electrodes) and translate these signals in (nearly) real-time into functional outputs that replace, restore, or augment natural brain capabilities (Wolpaw et al., 2002; Wolpaw & Wolpaw, 2012). This broad classification is recognized by the BCI Society and reproduces a consensus in neurotechnology scholarship: BMIs enable users to communicate or control devices directly through neural signals (BCI Society, 2024).

Secondly, and within the BMI context, Artificial Intelligence (AI) refers to advanced machine-learning models and deep neural networks that are used to process and identify intricate neural signal patterns: these AI models are able to decode user intentions into specific actions by detecting and deciphering motional or cognitive commands from noisy, high-dimensional neural data (Barredo Arrieta et al., 2019).

This chapter will, therefore, interrogate the overlapping ethical dimensions of the crossing of these technologies: how AI-enhanced BMIs challenge our traditional concepts of responsibility, autonomy, and safety. Some of the questions we will touch on will be the following: Can users of AI-BMIs be (truly) autonomous in cases where their actions emerge from technological mediation? Also, what are the boundaries of responsibility in cases when an AI model triggers particular bodily action? And how do we guarantee that these systems are safe, reliable, and transparent for both the users and experts?

To answer to these queries, we will advance with three main goals that will offer the kind of relevant normative analysis that we consider to be relevant to understand the ethical reflection of AI-BMIs: first, we will consider both invasive (e.g. intracortical implants for motor prostheses) and non-invasive BMI (e.g. scalp-based neurofeedback in locked-in syndrome patients) so we can contextualize, in the clinical practice, how AI can be deployed and used to achieve different concrete outcomes. Moreover, we will try to identify specific ethical tensions that can be raised by these BMI-AIs tools, discussing three ethical concerns:

(a) how the autonomy and agency of patients can be disrupted by the mediation of these technological developments (Vlek et al., 2014);

(b) how the physical safety of patients – particularly in the long run – , can be problematic due to the high risk of infection, inflammation and surgical complications that are hard to predict (Ryu and Shenoy, 2009);

(c) how the technological reliability can be guaranteed when faced with possibilities of "crosstalk", signal or biological degradation or issues of lack of transparency of many AI systems (Barrese et al., 2013).

Finally, we present a normative framework that will help us to support our ethical analysis, grounded in the idea that the technology used in AI-BMIs scenarios ought to be transparent for users, medical experts and everyone involved in the development and use of the tool. We will also propose an ethical oversight structure based on the idea of expert ethical boards, one that can emerge from the field of international Neuroethics (van Stuijvenberg et al., 2024) similar to what happened with Ethics Committee for hospitals and universities in the previous century.

Following this tripartite structure (mapping, evaluation, and normative framing) we will be able to create a relevant narrative that will take both empirical descriptions to moral analyses in a constant dialect in order to achieve some appropriate practical recommendations that, hopefully, may be considered by those developing these kinds of technological developments.

2. Technological Integration: AI and BMIs

2.1 Brief History and Current State of BMI Development

The first Brain–Machine Interface (BMI) development traces back to some of the foundational experiments in bioelectricity, by Hans Berger: in 1924, he produced the first human electroencephalogram (EEG), identifying both α- and β-wave patterns and signaling the dawn of neural signal measurement (Berger, 1929). Developing this and other methods, researchers were able to interfacing directly with brain activity: Delgado's invasive implants demonstrated neural stimulation in both animals and humans, and Vidal coined the term "brain–computer interface" in 1973 (demonstrating single-trial EEG control of cursors (Delgado, 1969; Vidal, 1973).

Based on these breakthroughs, Kennedy and Bakay were able to use the first human intracortical BMI implant in 1998 in a patient suffering of a locked-in syndrome (Kennedy & Bakay, 1998). Seven years after, a tetraplegic patient was able to use its own neural signals to control a robotic arm, implanting 96 electrodes (Hochberg et al., 2006). The same kind of achievement was made through non-invasive methods: for example, the SMR-based cursor navigation system developed in the early 1990 by Wolpaw in a successful case of BMI without any invasive component (Wolpaw et al., 2002).

In the last decade, the diversity and scale of BMIs has been raising with many different projects and applications: for example, DARPA-funded initiatives and private ventures (e.g., Neuralink, Kernel, Facebook) are now exploring ways to increase the viability and bidirectionality of both invasive and non-invasive methods of BMIs (DARPA, 2019): invasive methods are based on intracortical electrodes that are implant in the brain with high signal and spatial resolution, while the non-invasive methods (such as EEG, ECoG or NIRS) can provide less risk of harm and dependence in the long-run (Gosh, 2023).

Now, the goal of integrating Artificial Intelligence into BMIs can be conceptually confusing since it broads the functional boundaries of what is a traditional technology

The integration of Artificial Intelligence (AI) into Brain–Machine Interface (BMI) systems has redefined the conceptual and functional boundaries of neurotechnology: at first, since more rudimentary tasks were needed (e.g. cursor control, binary selection task, etc.), linear decoding models such as the Wiener and Kalman filters were more than enough (they assume a sort of stable relationships between the task (i.e. motor output) and the intention (i.e. neural inputs) (Kim et al. 2008). However, for more complex tasks, these models are very problematic when applied to high-dimensional neural data that are non-stationary by definition, which can be found in multi-task or long-term contexts.

Because of this, data-driven models using machine and deep learning had to be replace these deterministic models: the traditional methods (e.g. Linear Discriminant Analysis or the Support Vector Machines), even if they were quite effective in processing signals from noise than the previous methods, cannot achieve the kind of decoding accuracy, specificity (for different users) and generalizability that these new models can (Glaser et al., 2017).

The raising and applications of deep neural networks (e.g. Convolutional Neural Networks and Long Short-Term Memory) have shown to be more efficient in processing the complex spatial-temporal features that arise from different neural signals of complex task: these new models can identify the distribution of complete neural patterns and can, for example, decode speech from cortical activity, raise the ability of motor control or help users to reconstruct visual imagery (Hu et al., 2023).

This new development, the AI-BMIs, can allow for co-adaptive decoding methods that can adapt way more easily to the intra-subject variability that will naturally occur over time to a patient: by using online and reinforcement learning methods, this can help to face what is known as "neural drift", signal degradation and, more importantly, the electrode reconfiguration. If we consider patients with chronic implanted systems, this can allow achieving a long-term stability and adaptability that is very useful from a clinical point of view (both for the efficiency of the tool, but also for the well-being of the patient) (Orsborn, et al. 2012).

Some recent applications of this kind of AI-BMIs technology can be mentioned: for example, Nvidia partnered with Synchron by developing specific edge-AI modules (Holoscan platform by Nvidia) which enable low-latency decoding of wireless stentrode BMI (a less invasive method that implants a stent-like device not in the brain, but in a blood vessel near the brain) (Synchron, 2025). This improves both the technical difficulties present in more traditional methods, but also raises the commercial viability of these technologies, which can allow, in a near future, to diversify the applications and uses.

This novel integration of Artificial Intelligence into Brain–Machine Interface technology, even if it can bring many positive outcomes from a clinical point of view, raises many ethical and social queries that also need to go hand-a-hand with the technical development. For example, it is not clear at all if data privacy or signal interpretability can be fully guaranteed: the application of AI methods can develop specific algorithmic bias that were not found in more traditional versions of BMIs. As it is well-known, neural data are immensely personal and, therefore, sensitive from a privacy perspective: they can carry, at least in theory, the potential to reveal more than just the expected clinical outcome (e.g. a motor intention) but also, and perhaps even more, emotional, cognitive and thinking states (Zhou et al., 2022).

If we consider the problem of interpretability of many AI models that can be applied to BMIs, then we can have similar "AI-BMI" black box problem that we see currently happening with some deep learning architectures or generative transformers: since they are quite obscure – from an epistemic point of view –, this may raises consequences for the kind of trust users and experts can have in this technology, but also the in the kind of clinical validation that these methods should achieve to be widely used (Varshney and Alemzadeh, 2017).

Other ethical considerations that should be relevant to consider are the potential to – and again, at least in theory – allow cognitive or behavioral influence on those using the technology in question: this raises obvious concerns regarding traditional concepts of medical bioethics such as autonomy, informed consent or the integrity of personal identity (Yuste et al., 2017). To avoid missing the "ethical train", we will propose that specific multidisciplinary regulatory frameworks (one that considers the collaboration of different expertise) ought to be consider before the technology is widely available and used in clinical or personal contexts.

Less direct but also highly relevant are the socio-economic inequalities that can be worsened even more: on one hand, this technology is very costly and, therefore, it may be only accessible to those with economic power and, by consequence, inaccessible to those that cannot acquire these clinical tools (Rainey et al., 2020); on the other hand, the potential misused in other contexts – e.g. military, surveillance –, demands for careful oversight on how this technological development will be developed and deployed (Ienca, Jotterand, & Elger, 2018).

Next, we will analyze some of the key applications of this neuro-technology in clinical and other contexts to provide empirical contexts relevant for the ethical analysis.

2.2. Key Applications in AI-BMIs and a Case Study

As we noticed, Artificial Intelligence-driven Brain–Machine Interfaces (AI-BMIs) are starting to transform many different applications in several domains (e.g. cognitive, communicative or clinical). These new kinds of neuroprosthetics can directly translate neural signals into a wide range of commands which can be applied to control wheelchairs, prosthetic limbs or even full exoskeletons, restoring patients that were affected, for different reasons, by strokes, limb loss or paralysis (Wolpaw and Wolpaw, 2012).

The applications of AI can specifically enhance both the speed and accuracy of the BMI, which provides, from a user point of view, more natural and intuitive control. Moreover, AI models can help to increase the level of adaptability for both calibration and error correction of neural signals, which allows from a higher user comfort and durability in the long run (Horváth, Kovács & Benkő, 2024).

Psychologically, these developments also can affect the reintegration of patients using the technology, since they can be inserted in social and occupational activities that, before this use, could not happen.

If we consider cognitive enhancement, AI-BMIs can potentially augment different aspects of the human mind: with their capacity to real-time detection and constant modulation of neural patterns connected to working memory, executive function or attention, this can open the door to different kind of enhancement that can raise normative considerations since it can allow humans to perform beyond normal levels of performance (Tsai et al. 2025). The same technology can also be applied to increase the capacity of treatments in neuropsychiatric disorders: through AI-drive neurofeedback systems, patients with unbalanced brain functions (e.g. anxiety or ADHD) can receive closed-loop interventions (which are tailored to each individual's neural networking) with promising results (Sitaram et al., 2017). For healthy individuals, several non-invasive techniques (e.g. transcranial direct current stimulation or transcranial magnetic stimulation), merged with AI models, are currently being explored and tested to enhance both the learning and cognitive performance (Zhang et al. 2020).

A particularly innovative application of AI-BMI can actually be considered not a BMI (at least in the traditional sense): brain-to-brain communication (B2BC) is a new way to set up different mental agents (e.g. two persons, two monkeys, etc) and share relevant information in both to achieve a specific and mutual task: some experimental studies have been already performed using non-invasive EGG and transcranial stimulation, with the support of AI models to decode and encode neural signals to share commands, abstract information of images (Rao et al., 2014). Of course, this new technology, being a further development of more traditional BMIs, are accompanied by the same ethical problems identified before, such as concerns for privacy, autonomy and the influence on the self of those using the technologies. However, it also raises new,[2] original ethical problems, such as the problem of responsibility in collective decision-making (that is, when two people are interacting using a B2BC, who ought to be responsible by the outcomes of that use?) or in the raise of new kind of social interactions never seen before (Yoo et al., 2023).

An incredible case study of BMI influenced by AI is the pioneering research developed by Miguel Nicolelis and his collaborators. In one of the most emblematic moments for humanity, a paraplegic patient named Juliano Pinto was implanted with a BMI system with the goal of performing the first official kick ball of the World Cup 2014 in Brazil by controlling a full exoskeleton (BBC, 2014).

The BMI used was an invasive one, where many microelectrodes were implanted within Pinto's primary motor cortex focused on the leg movement: this allowed for a recording (in high-resolution) of action and field potentials, providing a complex map of data that recreated the motor intentions of the user. But this was not a "simple" BMI, but an AI-based technology that was able to decode and translate the complex neural patterns into concrete actions directed to the robotic exoskeleton: machine learning methods were used (from linear and non-linear to more adaptative techniques) in order to captured, on one hand, the dynamic nature of neural activity and, one the other hand, the non-stationary brain patterns, allowing the BMI to maintain relevant control of the technology even with high neural changeability

[2] B2BC also raises new technical challenges such as improving information bandwidth, reliability, and user training protocols which have indirect ethical concerns.

and signal drift (Nicolelis and Lebedev, 2009; Nicolelis and Lebedev, 2014).

To understand the complexity of the application of this kind of technological tool, it is interesting to describe some of the incredible process that Pinto's add to endure in order to be able to control, with its own mind, a complex exoskeleton. After the surgical implantation of the electrodes, the patient had to record his neural activity in both real and imagined motor movements during several weeks: the BMI design allowed to incorporated co-adaptative learning paradigms where decoding algorithms were used to optimize the performance. This bidirectional arrow is highly relevant to achieve the kind of natural fluidity to perform a soccer kick: of course, the complexity of the robotic exoskeleton was also important since it recreated the appropriate biomechanics for that particular movement (Colucci et al., 2022).

This technical achievement shows the potential of AI-BMIs for clinical use but also showed how invasive methods can be quite critical due to their surgical risks and long-term biocompatibility (in the case study used, it is not clear that Pinto will be able to access the exoskeleton for other more important tasks than kicking a soccer ball). Also, it showed how the perception of the nature of human disability can change with the use of technology broadly considered.

With this brief overview of the AI-BMI current paradigm, we will now focus on the ethical challenges that are raised by these technologies. We will provide a framework to evaluate, concretely, its normative limits and standards.

3. Ethical Concerns in AI-driven BMIs

3.1 Invasive vs. Non-Invasive BMI Methods (AI-Driven)

As previously mentioned, AI-BMIs can be divided into two main modalities: invasive and non-invasive methods. Each of them represents different clinical and technical details and, more importantly, specific ethical considerations that ought to be made.

Invasive BMIs usually involve those applications that require a direct implantation, in the cortical surface or the parenchyma of the brain, of electors that enable to acquire and replicate very precise neural signals to control external devices: the signals have to be decoded and interpreted by AI algorithms which excel at identify temporal and spatial patterns from the brain activity (Jiang et al., 2023).

Differently, non-invasive BMIs mainly depend on external sensors such as functional near-infrared spectroscopy (fNIR), magnetoence-phalography (MEG) or, more famously, electroencephalography (EEG), which enables to monitoring of the neural activity not via the brain, but the scalp or skull. The idea here is to avoid the physical risks that any brain surgery has (more soon): but to avoid this, something else is lost, namely, a lower spatial resolution in conjunction to higher levels of "noise", which influence the efficiency of the technology. Here as well AI-driven methods have been relevant to enhance these two problems: by applying specific models, non-invasive methods can increase the classification accuracy, feature extraction and improve the signal decoding (Zhao et al., 2021).

Some of these techniques are RNNs (recurrent neural networks), SVM (support vector machines) or CNNs (convolutional neural networks), which can increase the interpretation of the non-invasive signals and, therefore, rise the overall accuracy of the BMI in use.

If we take into consideration the clinical perspective, BMIs have been quite useful regarding the performance of fine motor control or multi-degree-of-freedom commands (e.g. speech decoding, robotic arm/leg movement, etc,) (Hu et al., 2023). Invasive BMIs are more interesting in this task, since they can take into account the neural plasticity and the signal non-stationarity, which allows to avoid changes of performance over time. Nevertheless, as we will see below, invasive BMIs have high surgical risks and biocompatibility limits that should be ethically considered, especially if we take into consideration the widespread application of this technology for different populations. This is where non-invasive BMIs can have an important role: if we look at safety and ease of use, this kind of technological application seems better to monitoring cognitive states, to help locked-in-patients communicate or to provide neurofeedback therapies (Krigolson et al., 2020).

From a normative perspective, the decision to embrace invasive vs non-invasive methods raise pertinent questions about patient autonomy and well-being, trade-offs benefit between harm and efficiency, and the kind of informed consent that should be based on known factors that can be predicted over time. If it is true that invasive BMIs offer high control capabilities, the long-term and neural impact of implementing electrodes in the brain of patients ought to be clearly indicated; on the other side, non-invasive BMIs are safer (when compared to the invasive ones), but they face high limitations regarding the functionality and reliability of the technology, which can affect in some cases the autonomy of users (e.g. in cases where the technology does not perform what the user is intending) or the dignity of users (e.g. when a technology is been used but does not bring an practical relevant advantage).

Next, let’s detail some of the risks that can occur in Invasive BMIs.

3.2. Risks of Surgical Intervention in BMIs

As we saw before, invasive BMIs are interesting from the efficiency point, but it introduces many different medical risks that can have both clinical and ethical considerations. Let us consider some of the stages required to implement a common BMI system, such as intracortical implants.

First of all, it should be noticed that these kinds of procedures are very expensive from a resource perspective: they require a coordination between engineering, neuroscientific and clinical areas that develop, throughout a timeline that can extend from several months to years (depending of the complexity and the degree of motor control intended), the entire system.[3]

The process starts with a pre-surgical phase that is non-invasive and is focused on determining the cortical targets of the patients: this involves high-resolution brain imaging such as fMRI, MRI or CT crossed with neuropsychological data (typically from the primary cortex motor when focused on motor BMI) (Flesher et al., 2021). In this stage, informed consent measures are taking place even if the long-term implantation are highly correlated with unknow risks, which creates a very delicate ethical situation in these cases.

The next phase is the actual surgical implantation of the electrodes: this happens in full anesthesia which can take up to 6 hours to conclude, depending of the kind of device chosen (e.g. ECoG grids, Stentrodes, etc). The post-surgical recovery can vary in time depending of the age and fitness of the patient, but it typically takes 2

[3] Neurosurgeons, biomedical engineers, neuroethicists, rehabilitation specialists, and AI developers are all involved in this process. Clinical trials often report resource use exceeding $100,000 USD per participant per year (Daly and Wolpaw, 2008).

weeks; after, the first signal quality testing happens: for example, Neuralink use 96-channel microelectrode to interface with external devices that decode the signals to achieve, over the course of many weeks, the kind of calibration needed. This takes daily sessions of training between 1-3 hours where a specific task is made (e.g. imaging an image and record the neural activity) (Orsborn et al., 2012).

In this stage, machine learning models (e.g. SVMs or recurrent neural networks) are being trained to identify neural patterns and cross those with the motor intentions. For that, it is needed at least 40 hours of recorded data that are constantly refined to achieve the best results possible (Collinger et al., 2013). Studies on full functional control of robotic limbs showed the need of intensive training between 3 to 6 months, depending of both the cortical baseline plasticity of the patient and the kind of task performed. During this time, many things can happen, from signal drifting to electrode degradation, which require a laborious and demanding constant technical work.

If we go back to the initial story of the 2015 World Cup kick, Juliano Pinto had to train for months not only the neural phase, but also the linking with the effector system (in this case, the exoskeleton), calibrating it to match the natural velocity, biomechanics and trajectories of the patient.

Looking at the different stages, we can see how much risk is embedded in this kind of procedures: infections and tissue damages can cause severe complications and neurological discrepancies, at least in the long-term. Since the electrodes have to breach the protection of the skull and meninges, they can easily become a source of microbial infection, which can create all kind of clinical problems (e.g. meningitis, sepsis, abscess, etc.) which may need aggressive treatments or even, in worse cases, the removal of the implant, exposing the patient to even more risks while losing all the potential therapeutic benefits (Wolpaw and Wolpaw, 2012). It should also be clear that the insertion of electrodes disrupts neural tissues which can cause injuries to several neurons (and glial cells) due to the hemorrhage, causing inflammatory responses that can lead to further neuronal loss.

The motion between the actual brain tissue of the patients and the rigid electrode during the normal use of the device, since they are biologically different, can cause many tissues trauma and chronic tissue stress, which explains why there is a signal degradation over time. Moreover, the glial scarring can also cause negative effects on surrounding brain structures, which can lead to more neural degeneration and can affect the overall efficiency of the technology, which require a biocompability between the implant and the brain tissue that is hard to achieve currently (Otte, et al., 2022).

If we take all these empirical details into consideration, we can easily see how the surgical implantation causes some ethical reserves in the risk-benefit calculus of invasive BMIs: on one hand, the principle of non-maleficence requires diminishing harm and therapeutic gain if harm is done; but it is not clear that these can be guaranteed by the multidisciplinary team, even if informed consents may include an exhaustive list of the risks. We also know that many of these patients are in a fragile epistemic state due to their unhealthy situation and, therefore, are more likely to accept risks that they would never accept if healthy.

Therefore, we can claim that both infection and tissue damages, in the long run, are causally associated with this method and, therefore, we have a moral need to consider adequate responses to these long-term negative effects: addressing them will require better science, that is, will require design improvement, developing more flexible and biocompatible electrodes that can avoid infections and damage in the implantation area (Kozai et al., 2015).

3.3. Autonomy and Responsibility in BMIs

When we think in traditional concepts applied to healthy individuals, the notion of autonomy and responsibility are more or less easily applied if you ignore the metaphysical debates on these two tricky concepts. But when we apply those standards definitions in context of BMIs, things may get philosophically obscure since these concepts are related to conscious intentionality and control over the movements of our bodies. Since BMIs can act exactly on that (i.e. conscious and movement), they introduce new modes of interaction and, by consequence, of relevant impact in the world. Therefore, we can ask: what can count as a voluntary act of the user when using an interface that modulates the interaction between is conscious intentions and the specific execution of a movement? (Mele, 2009)

Firstly, if we take into consideration the fact that some of the neural signals that control a movement may not fully reflect a conscious intention (in the full sense), particularly if you take into consideration involuntary fluctuations in different areas of the brain that can be wrongly decoded, this can result in unintended performances from the user (Wolpaw and Wolpaw, 2012). Second, even if training and mental strategies can help to calibrate the control of the device, latency and errors can always occur, transforming control in a probabilistic system rather than a deterministic one. But if it is probabilistic, then a direct causal relationship between actions and intentions cannot be made and, therefore, this undermines the traditional concepts of responsibility we use (Hochberg et al., 2012).

Finally, when we use semi-autonomous (or even fully autonomous) AI models to decode or optimize signal performance, then this becomes even more problematic: we are dealing with a kind of shared agency that is not ethically considered in our legal and moral codes. Consequently, there is an urgent need to reconceptualize our concepts of autonomy and responsibility to one that takes into consideration a distributed perspective that is more relational than binary, that can provide a clear spectrum of fully intentional to partial or unintended actions (Gouveia, 2019).

This will demand an upgrade from intentional ethics to a kind of non-intentional actions, where human agents are no longer the only players in town: hybrid agent systems are arising and we need to develop relevant ethical frameworks to deal with them. This is especially relevant when we have some relevant empirical research that show that BMIs create, often, an illusion of agency, when users consider that an action is their initiative (Vlek et al., 2014). This disruption of agency challenges how users using BMIs perceive their own actions and the authenticity of their acting: there is a clear dissociation between what is the subjective experience of an intention and the objective control of that intention. If you take, for example, into consideration the delay of signal processing in some contexts, this temporal disruption can clearly affect the agency of the user, even if they perceive it as a legitimate act (Haggard, 2017).

These challenges require a level of care and attention by BMI experts in the training phase that may ensure that patients are aware of the kind of limits and alienation phenomena that may occur under the use of a BMI, in order to respect its autonomy but also its well-being in the long run.

3.4. Technical Challenges of BMIs

To conclude, we review some of the most fundamental technical challenges of BMIs that should be taken into accounting when thinking the morality of these technological applications.

First, as it was mentioned before, there are challenges regarding the signal reliability that can be achieved with AI-powered BMIs: even if it is true that this new version works better than traditional BMIs, the quality and stability of brain signals over time affects directly the performance which, in the long run, poses a significant challenge. The signal degradation can also occur with micromovements of the implant that can cause tissue reactions that reduce the functioning ratios (signal-to-noise) which affect, by consequence, the mental representation relevant for a specific task (Campbell and Wu, 2018). Non-invasive methods solve some of these problems, but they offer a low spatial resolution and can be easily interfered by external sources, which affects their reliability in the short time.

We can, however, claim that what is needed is actually better AI tools that can recalibrate, over time, the signal fluctuations, enabling a better performance even if degradation occur. Nevertheless, the use of these kinds of AI models introduces their own philosophical problems: specifically, many of these models have a black-box structure, which raises transparency issues that challenge the autonomy of patients in general (Gouveia, 2025). They can lead to algorithmic bias or difficulties in detecting failures which cannot be identified and, by consequence, corrected. Explainable AI frameworks can be an alternative, but these methods are still being explored in the present (Gouveia & Malík, 2024)

A second issue that is raised is the possibility of crosstalk or interference of the functioning of the implanted electrode. The typical invasive methods rely on multi-electrode arrays which can affect the individual performance of each electrode by unintentionally overlap or mix different signals of the adjacent electrodes, degrading the relevant neural signals for a specific task (Dunlap et al., 2020). There are several factors that can cause this phenomenon: resistance incongruity, neural plasticity, electrical noise and the physical proximity of each electrode, all of those affecting the quality of the signals recorded, which can lead to an inaccurate outcome for the user (Nicolelis and Lebedev, 2009).

To solve this issue, experts are trying to focus on improving the hardware design of the electrodes with enhancing the techniques for signal processing: this may involve, on one hand, improving the materiality[4] of which the electrodes are made of and their spatial distribution in the brain and, on the other hand, develop deep learning techniques to blind each electrode's activity and avoid the overlapping of signals to improve the decode precision (Liu et al., 2024)

From an ethical point of view, the possibility of crosstalk raises higher demands that we currently have in the field, since they affect both the efficiency but also the safety of users using an AI-BMI, which can lead to loss of trust in the technology (and psychological frustration or harm) or, in worse scenarios, actual physical harm. Therefore, we can claim that ensuring the best signal performance possible is not only a technical goal but actually, a normative standard that should be in place in every single team that develop BMIs applications.

[4] Flexible, ultrathin, and bioactive-coated can reduce scar formation and extend recording stability.

4. Conclusion: Some Ethical Tools

We saw in this chapter how AI-driven Brain-Machine-Interfaces raise many interesting clinical applications but also some ethical queries that we ought to consider. To deal with them, a couple of traditional ethical perspectives should be in place.

First, there is a need to take seriously both ethical review boards or committees and regulatory frameworks adapted to the specific details of this technology: given the impactful nature of BMIs in the mental life of users and the potential risks associated (e.g. mental privacy, disruption of personal identity, impact on the control and agency, etc) (Ienca and Andorno, 2017). Therefore, we need regulations that focus not only in the physical/clinical impact, but also in the mental "freedom" and respect of personal autonomy of users.

Ethics committees should also have a role here in applying these regulatory frameworks: they should be multidisciplinary by nature (in order to incorporate a holistic perspective of a human being), and they should act not only in the approval phase, but constantly throughout the different stages of the development and use of the technology. At the same time, there is an urgent need for what can be known as "international" harmonization of these different regulation and committee roles to be sure to avoid at all cost "ethics dumping" (Floridi et al., 2018). Finally, create an ethical culture and environment within the institutions and industries that develop AI-BMIs should be mandatory, where ethical reflection is seen as an essential feature of the work produced instead of a mere secondary aspect.

In sum, the amazing clinical potential of AI-BMIs necessitates a sustained ethical approach that can soundly informed policy-making and regulation while holistically approach neurotechnologies having in human relevant societal values that ought to be always present and never forgotten.

References

BBC News (2014), 'Paraplegic Man Controls Robot to Kick Off Brazil World Cup', 12 June, Available at: https://www.bbc.com/news/technology-27714214 (Accessed: 13 June 2025).

Borton, D.A., Yin, M., Aceros, J. & Nurmikko, A. (2013) "An implantable wireless neural interface for recording cortical circuit dynamics in moving primates", *Journal of Neural Engineering*, 10(2): 026010.

Campbell, A., & Wu, C. (2018) "Chronically Implanted Intracranial Electrodes: Tissue Reaction and Electrical Changes", *Micromachines*, 9(9): 430.

Clausen, J. (2011) "Conceptual and ethical issues with brain-hardware interfaces", *Current Opinion in Psychiatry*, 24(6): 495–501.

Collinger, J., Wodlinger, B., Downey, J., Wang, W., Tyler-Kabara, E., Weber, D., McMorland, A., Velliste, M., Boninger, M. & Schwartz, A. (2013) "High-performance neuroprosthetic control by an individual with tetraplegia", *The Lancet*, 381(9866): 557–564.

Colucci, A., Vermehren, M., Cavallo, A., Angerhöfer, C., Peekhaus, N., Zollo, L., Kim, W., Paik, N., & Soekadar, S. (2022) "Brain-Computer Interface-Controlled Exoskeletons in Clinical Neurorehabilitation: Ready or Not?", *Neurorehabilitation and neural repair*, 36(12): 747–756.

Daly, J. and Wolpaw, J. (2008) "Brain-computer interfaces in neurological rehabilitation", *The Lancet Neurology*, 7(11): 1032–1043.

Dunlap, C., Colachis, S., Meyers, E., Bockbrader, M. and Friedenberg, D. (2020) "Classifying Intracortical Brain-Machine Interface Signal

Disruptions Based on System Performance and Applicable Compensatory Strategies: A Review", *Frontiers in Neurorobotics.* 14:558987. doi: 10.3389/fnbot.2020.558987.

Fan, C., Hahn, N., Kamdar, F., Avansino, D., Wilson, G. H., Hochberg, L., Shenoy, K. V., Henderson, J. M., & Willett, F. R. (2023). Plug-and-Play Stability for Intracortical Brain-Computer Interfaces: A One-Year Demonstration of Seamless Brain-to-Text Communication. *Advances in neural information processing systems*, 36, 42258–42270.

Flesher, S., Collinger, J., Foldes, S., Weiss, J., Downey, J., Tyler-Kabara, E., Bensmaia, S., Schwartz, A., Boninger, M. and Gaunt, R. (2021) "Intracortical microstimulation of human somatosensory cortex", *Science Translational Medicine*, 13(598): eabe0662.

Floridi, L., et al. (2018) "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations", *Minds and Machines*, 28(4): 689–707.

Gouveia, S. (2019) "A Defense of the Principle of Distributed Responsibility in Artificial Intelligence", *Revue Roumaine de Philosophie*, 63(2): 235–248.

Gouveia, S. (2025) "The Ethics of Artificial Intelligence in Medicine: Preliminary Remarks", *Global Philosophy*, 35, 4. https://doi.org/10.1007/s10516-024-09737-y.

Gouveia, S. and Malík, J. (2024) "Crossing the Trust Gap in Medical AI: Building an Abductive Bridge for xAI", *Philosophy and Technology*, 37, 105. https://doi.org/10.1007/s13347-024-00790-4.

Haggard, P. (2017) "Sense of agency in the human brain"; *Nature Reviews Neuroscience*, 18(4): 196–207.

Hochberg, L., Bacher, D., Jarosiewicz, B., Masse, N., Simeral, J., Vogel, J., Haddadin, S., Liu, J., Cash, S., van der Smagt, P., and Donoghue, J. (2012) "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm", *Nature*, 485(7398): 372–375.

Horváth, Z., Kovács, A., & Benkő, Z. (2024) "Enhancing motor imagery classification in brain–computer interfaces using deep learning and continuous wavelet transform", *Applied Sciences*, 14(19): 8828.

Hu, Y., Zhang, Y., Zhang, S., Wang, X. and Sun, H. (2023) "EEG-based emotion recognition using hybrid CNN–LSTM with self-attention", *Biomedical Signal Processing and Control*, 82, p.104589.

Ienca, M., Andorno, R. (2017) "Towards new human rights in the age of neuroscience and neurotechnology", *Life Sci Soc Policy*, 13, 5. https://doi.org/10.1186/s40504-017-0050-1.

Ienca, M., Jotterand, F., & Elger, B. S. (2018) "From Healthcare to Warfare and Reverse: How Should We Regulate Dual-Use Neurotechnology?", *Neuron*, 97(2): 269–274.

Kozai, T., Jaquins-Gerstl, A., Vazquez, A., Michael, A. and Cui, X. (2015) "Brain tissue responses to neural implants impact signal sensitivity and intervention strategies", *Annual Review of Biomedical Engineering*, 17: 45–72.

Koralek, A., Jin, X., Long, J., Costa, R., & Carmena, J. (2012) "Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills", *Nature*, 483(7389): 331–335.

Krigolson, O., Williams, C., Norton, A., Hassall, C., and Colino, F. (2020) "Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research", *Frontiers in Neuroscience*, 14: 1–14.

Liu, X., Gong, Y., Jiang, Z., Stevens, T. and Li, W. (2024) "Flexible high-density microelectrode arrays for closed-loop brain–machine interfaces: a review", Frontiers in Neuroscience, 18:1348434. doi: 10.3389/fnins.2024.1348434.

Mele, A. (2009) *Effective Intentions: The Power of Conscious Will*, New York: Oxford University Press.

Nicolelis, M. and Lebedev, M. (2009) "Principles of neural ensemble physiology underlying the operation of brain–machine interfaces", *Nature Reviews Neuroscience*, 10(7): 530–540.

Nicolelis, M. and Lebedev, M. (2014) "Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation", *Physiological Reviews*, 97(2): 767–837.

Orsborn, A.L., Dangi, S., Moorman, H.G. and Carmena, J.M. (2012) "Closed-loop decoder adaptation on intermediate time scales facilitates rapid BMI performance improvements independent of decoder initialization conditions", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(4): 468–477.

Otte, E., Vlachos, A. & Asplund, M. (2022) "Engineering strategies towards overcoming bleeding and glial scar formation around neural probes", *Cell and Tissue Research*, 387(3): 461–477.

Rainey, S., Martin, S., Christen, A., Mégevand, P., & Fourneret, E. (2020) "Brain Recording, Mind-Reading, and Neurotechnology: Ethical Issues from Consumer Devices to Brain-Based Speech Decoding", *Science and engineering ethics*, 26(4): 2295–2311.

Rao, R., Stocco, A., Bryan, M., Sarma, D., Youngquist, T., Wu, J., and Prat, C. (2014), "A direct brain-to-brain interface in humans", *PLOS ONE*, 9(11), e111332.

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Rana, M., Oblak, E., Birbaumer, N. & Sulzer, J. (2017) "Closed-loop brain training: the science of neurofeedback", *Nature Reviews Neuroscience*, 18 (2): 86–100.

Tsai, P, Akpan, A., Tang, K. and Lakany, H. (2025) "Brain computer interfaces for cognitive enhancement in older people - challenges and applications: a systematic review", *BMC Geriatrics*, 25, 36..https://doi.org/10.1186/s12877-025-05676-4 van Stuijvenberg O.C., Samlal DPS, Vansteensel MJ, Broekman MLD and Jongsma KR (2024b) "The ethical significance of user-control in AI-driven speech-BCIs: a narrative review", *Frontiers in Human Neuroscience*, 18: 1420334. doi: 10.3389/fnhum.2024.1420334.

Vlek R, Van Acken J, Beursken E, Roijendijk L, Haselager P. (2014) "BCI and a user's judgment of agency". In: Grubler, G and Hildt, E. (eds.) *Brain-computer-interfaces in their ethical, social and cultural contexts*, Dordrecht: Springer, pp. 193–202.

Yoo, S. H., Choi, K., Nam, S., Yoon, E. K., Sohn, J. W., Oh, B. M., Shim, J., & Choi, M. Y. (2023) "Development of Korea Neuroethics Guidelines", *Journal of Korean Medical Science*, 38(25): e193.

Yuste, R., Goering, S., Arcas, B. A. Y., Bi, G., Carmena, J. M., Carter, A., Fins, J. J., Friesen, P., Gallant, J., Huggins, J. E., Illes, J., Kellmeyer, P., Klein, E., Marblestone, A., Mitchell, C., Parens, E., Pham, M., Rubel, A., Sadato, N., Sullivan, L. S., … Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, *551*(7679), 159–163.

Wolpaw, J.R. and Wolpaw, E. W. (2012) *Brain–computer Interfaces: Principles and Practice*, New York: Oxford University Press.

Zhao, Z., Zhang, H., Liu, Y., and Chen, X. (2021), 'Deep learning-based decoding of complex motor imagery for neuroprosthetic control', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29: 250–259.

Zhang, X., Ma, Z., Zheng, H., Li, T., Chen, K., Wang, X., Liu, C., Xu, L., Wu, X., Lin, D., & Lin, H. (2020) "The combination of brain-computer interfaces and artificial intelligence: applications and challenges", *Annals of Translational Medicine*, 8(11): 712.