Human Dignity and the Future of Personhood in AI-Integrated Brain-Computer Interfaces.

Isaac Mutelo (PhD)
Lecturer & Director of Quality Assurance/Research, Innovation and Industrialisation,
Arrupe Jesuit University, Harare, Zimbabwe

Mohammed X Ntshangase (PhD)
Lecturer of Philosophy of Education/ Psychology,
University of Limpopo (Turf. Campus), South Africa

Godfrey Gotora (MEng)
Lecturer, School of Engineering and Information and Communication Technology,
Arrupe Jesuit University, Harare, Zimbabwe

Abstract

The advancements in neurotechnology, especially brain-computer interface (BCI) technologies, signal an ethically challenging phase in the relationship between human beings and artificial intelligence. The boundaries between 'purely biological cognition' and machine intelligence continue to blur. The current objectives of tech giants such as Neuralink are primarily therapeutic, focusing on restoring communication and neurological functions. However, future developments may enable direct integration between the human brain and AI-integrated BCIs such as generative artificial intelligence (GenAI) systems through implantable brain chips or other devices. This will create new possibilities for cognitive enhancement, expanded memory, accelerated learning, and augmented decision-making. This conceptual paper critically examines the ethical implications of AI-integrated brain enhancement from the perspective of human personhood and dignity. Although the ethical implications of brain-computer interfaces have received growing scholarly attention, much of the existing literature has concentrated on issues such as safety, privacy, cybersecurity, regulation, and informed consent. Comparatively less philosophical attention has been devoted to examining AI-integrated cognitive enhancement through the combined lenses of personhood, human dignity, and relational understandings of the human person. The paper argues that the shift from therapeutic neurotechnology to human enhancement raises fundamental questions about the nature of human agency, autonomy, identity, and responsibility in an age where cognitive processes may increasingly involve technological mediation. It is crucial to consider whether AI-enhanced cognition represents an authentic extension of human capacities or a transformation that risks undermining essential dimensions of human personhood. Will AI-integrated brain enhancement support the conventional understanding of human personhood, or will it transform the very meaning of being human? Drawing on existing literature in technoethics, neurotechnology, and philosophy of technology, the study argues that the ethical evaluation of BCI technologies should not be determined solely by technological capability or efficiency, but by their capacity to preserve human dignity, meaningful autonomy, and responsible self-determination. As humanity moves towards deeper integration with artificial intelligence, the central ethical challenge lies beyond whether humans can enhance themselves through neural technologies. It is primarily centred on interrogating whether such enhancement can be pursued in ways that safeguard the distinctive value, freedom, and dignity of the human person. By integrating the extended mind hypothesis, the transformation hypothesis, human dignity, and *Ubuntu*, the paper proposes a normative framework for evaluating future AI-integrated neurotechnologies.

Keywords: Neuralink, brain–computer interfaces, generative artificial intelligence, personhood, enhancement, human dignity, autonomy.

Introduction

Brain-computer interface (BCI) technologies mark a crucial step in the evolving relationship between human beings and artificial intelligence. Wander and Rao [2014] and Mouazen and El Bouyed [2025] understand brain-computer interface as any technology that enables the human brain to communicate directly with a computer or other electronic device by converting neural signals into actionable commands. Current BCIs such as those being developed by the Neuralink Corporation are primarily used to help people with neurological disorders or physical disabilities, such as paralysis, spinal cord injuries, stroke, or amyotrophic lateral sclerosis (ALS), regain communication, mobility, or independence [Wander and Rao 2014]. For example, Neuralink Corporation [2026], an American neurotechnology company developing implantable brain-computer interfaces, notes that:

> *Our brain-computer interface translates neural signals into actions. In our clinical trials, people are using Neuralink devices to control computers and robotic arms with their thoughts. This technology will restore autonomy to those with unmet medical needs and unlock new dimensions of human potential.*

Current ambitions of brain-computer interface technologies such as Neuralink are primarily therapeutic, as they aim at restoring communication and neurological function for people with paralysis or severe disability [Reuters 2023; Mchia 2023; Neuralink Corporation 2026]. However, the longer-term vision may include the possibility that such devices may in the near future, connect the human brain directly to future AI-integrated BCIs. This will enhance cognitive abilities, expand memory, accelerate learning, and impart AI-assisted creativity and decision-making. Beyond technical and medical concerns, this shift raises fundamental philosophical questions regarding what it means to be human. For example, the shift challenges the traditional or conventional definitions of human dignity, autonomy, personhood, privacy, and social justice.

Regarding the computer-human interfaces, Livanis et al. [2024] argue that among the central ethical issues are "human dignity, personhood and autonomy," along with safety, stigma, discrimination, privacy, security, and responsibility. Such concerns become even more pressing when BCI technology moves beyond restoration toward enhancement. For instance, if neural devices are used not merely to compensate for disability or paralysis but to augment ordinary cognition, then the question is no longer only what humans can do with technology, but what humans may become. In that sense, the ethical significance of BCIs such as future implantable neural interfaces lies in their potential to redefine human personhood itself. AI integration can either be perceived as a continuation of human evolution and an expansion of human capacities (extended mind hypothesis) or as a process that alters the very definition of being human (transformation hypothesis). These perspectives should not be understood as mutually exclusive but as complementary analytical frameworks for evaluating different dimensions of AI-integrated neurotechnology. The tension is apparent when BCIs are described as technologies that connect "the human brain to a machine that is external to the brain and the source of the self," thereby implicating not just function but the very definition of selfhood [Mchia 2023:982].

In the age where thought itself becomes increasingly technologically mediated, it is crucial to rethink traditional accounts of autonomy and moral responsibility. Based on the highlighted dilemma, this paper interrogates whether AI-integrated brain enhancement will extend human personhood or transform it. From the perspective of contemporary neuroethics, this paper does not simply focus on safety or efficacy, but on the need to rethink how meaningful autonomy and human dignity can be preserved. In 2021, UNESCO warned that neurotechnology innovations must be approached with careful attention to human rights and ethical governance. For instance, a purely instrumental account of BCIs such as future implantable neural interfaces that measures success only by speed, efficiency, or intelligence may not adequately address deeper philosophical questions. Accordingly, this paper advances three related arguments. Firstly, AI-integrated BCIs can plausibly be understood as an extension of humanity's long history of technological self-enhancement. Secondly, the increasing integration of artificial intelligence into human cognition simultaneously creates genuine risks of transforming personhood, agency, and authenticity. Thirdly, because neither technological optimism nor technological pessimism adequately resolves these tensions, human dignity, understood within both universal human rights discourse and the African relational philosophy of *Ubuntu*, provides the most appropriate ethical framework for evaluating future developments in neurotechnology [Igbalumun & Mutelo

2025]. Given the rate at which BCIs are developing, the question is no longer whether humans can enhance themselves through neural technologies. Rather, the main question is whether such enhancement can be pursued in ways that safeguard the distinctive value, freedom, and dignity of the human person.

1. Extended Mind Hypothesis and AI-Augmented Intelligence

The main question regards whether AI-integrated enhancement actually represents an extension of human personhood. Is AI-integrated enhancement a continuation of humanity's long history of extending its capacities through technology? The extended mind hypothesis, originally formulated by Clark and Chalmers (1998), suggests that cognition may extend beyond the biological brain into appropriately integrated external artefacts. This is because human cognition is not confined to the physical brain but actively incorporates external tools [Garza-Herrera 2024; Mihić Zidar et al. 2026]. Historically, human beings have always operated with specific 'tools' that have shaped the very definition of thinking, remembering, creating and calculating. Garza-Herrera [2024] notes that "One of the defining characteristics of human evolution is the use of instruments. From the earliest stone tools to the most advanced artificial intelligence (AI)-powered devices, tools have assisted humans in surviving, thriving, and shaping the world around us."

In that sense, human beings have never been 'purely biological' in the sense of operating without the support of any form of tool. The fact that brain–computer interfaces "enable direct communication between the brain and external computers" makes them a new stage in the age-old pattern of technological augmentation [Gordon and Seth 2024]. In that sense, AI-enabled neurotechnologies are not an unexpected intervention in human history since it echoes the familiar use of artefacts to transcend biological limitations. For example, writing does not merely record thought but extends external memory beyond the fragile limits of the individual brain. Similarly, the internet extends human access to knowledge, enabling one to retrieve, compare, and distribute information at a scale no mind could manage. In their work on the digital expansion of the mind, Marsh and Rajaram [2019] and Yamashiro and Roediger III [2019] argue that internet use changes how people remember and process information. Their position demonstrates how cognition is already deeply entangled with external systems of storage and retrieval.

Just as books expand memory and the internet expands access to knowledge, AI-linked brain implants can be perceived as the next step in this continuum. It permeates the boundary between internal cognition and external assistance through enhancement. From this perspective, the question shifts from whether technology affects human cognition to whether AI-integrated enhancement simply intensifies a long-standing pattern of human self-extension [Gilbert et al. 2023; Mchia 2023]. If writing, books, the internet, and smartphones are cognitive partners, then linking the human brain directly to future generations of AI-enabled neural interfaces such as implanted brain chips may be seen as a more intimate version of the same phenomenon. Based on the extended mind hypothesis, BCIs would be seen as supplementing human intelligence just like calculators extend arithmetic ability without abolishing human reasoning. Gruica [2025:89] recognises that these systems are increasingly about more than medical restoration, since they may also involve "the potential for human enhancement and human-AI symbiosis." Thus, the main question is whether such augmentation enhances human agency or merely redistributes the way it is exercised. What if instead of imagining future generations of AI-enabled neural interfaces as a competitor to human thought, we view it as a tool for co-thinking that helps with recalling, pattern recognition, creative brainstorming, and providing decision-making support?

From the perspective of human–AI collaboration, Garcia-Barrios [2025:7] questions whether intelligent systems are "epistemic partners or cognitive crutches," emphasising that the long-term value of AI depends on whether it supports critical thinking and ethical agency rather than displacing them. Applied to AI-enabled neurotechnologies, this entails that AI should be understood as a partner in cognition rather than a replacement for the human mind [Pugh et al. 2018]. It can function as an external resource that enlarges memory, expands access to information, and accelerates learning without necessarily overriding human interpretation, judgment, and responsibility. But is technology that enhances reasoning, creativity, and memory simply expanding existing human potential? Based on the extended mind hypothesis, the AI-integrated brain

enhancement is only a more sophisticated instance of human development. If BCI technologies can restore lost capacities, disabilities or paralysis, they can also plausibly augment existing ones [Garcia-Barrios 2025; Garza-Herrera 2024; Mihić Zidar et al. 2026].

In that sense, enhancement should be regarded as a technologically mediated expression of human striving toward fuller realisation rather than loss of humanity. However, this optimistic reading of BCIs from the perspective of the extended mind hypothesis requires caution because technological extension cannot always be regarded as neutral. This is because every enhancement reshapes what users become accustomed to doing for themselves and what they outsource to machines. At the same time, the foregoing discussion emphasised that human personhood has always been dynamic, relational, and tool-mediated [Garza-Herrera 2024; Mihić Zidar et al. 2026]. On that basis, AI-integrated brain enhancement may be regarded as an extension of human personhood. This is partly because technology has historically functioned as a means through which humans transcend biological limitations and reimagine the scope of their own capacities. The table below demonstrates the progression.

Stage	Development	Implication
Technological extension	Tools such as writing, books and the internet extend human cognitive capacities.	Cognition is not limited to the biological brain.
AI augmentation	AI and BCIs extend memory, reasoning, creativity and access to information.	Human intelligence can be technologically enhanced.
Human–AI collaboration	AI functions as a cognitive partner rather than simply a tool.	Agency may be redistributed rather than eliminated.
Cognitive dependence	Increasing reliance on AI may shift what humans retain or outsource to machines.	Technological extension is not necessarily neutral.
Extension of personhood	AI-integrated BCIs represent a more intimate form of human technological self-extension.	Personhood may be extended without necessarily being transformed.
Theoretical limit	The extended mind explains cognitive extension but not deeper changes to identity and selfhood.	The transformation hypothesis is needed to examine possible ontological change.

While the extended mind hypothesis clearly demonstrates that human cognition has historically evolved through technological mediation, it cannot by itself resolve the deeper philosophical concerns raised by direct AI integration into human cognition. The hypothesis explains how technologies extend cognitive capacities without responding adequately to whether increasingly intimate technological integration might alter identity, authenticity, or the very meaning of personhood. These unresolved questions require a complementary perspective such as the transformation hypothesis, which is capable of examining not merely cognitive extension but possible ontological transformation.

2. Transformation Hypothesis and Personhood

The transformation hypothesis provides a progressive account of how increasingly intimate AI-human integration may alter conventional understandings of personhood [Goering et al. 2021; Kögel et al. 2019]. The evolution can be understood across several stages, beginning with cognitive enhancement and potentially culminating in a technologically mediated or co-constituted conception of the human person, as summarised in the table below.

Stage	Transformation	Implication for Personhood
Cognitive enhancement	AI enhances memory, reasoning and creativity.	The person becomes cognitively augmented.
Cognitive dependence	Human thinking increasingly relies on AI.	Autonomy may become uncertain.
Identity and authenticity	AI-mediated cognition affects memory, intentions and decisions.	The boundaries of self and authentic thought become blurred.
Privacy and agency	Neural data and mental processes become technologically accessible.	Mental privacy and agency become vulnerable.
Technologically mediated personhood	Human and AI cognition become deeply integrated.	Personhood may become hybrid and technologically co-constituted.
Transformation	Technology reshapes cognition, identity and self-understanding.	The question shifts from *what technology does to persons* to *what kind of person emerges*.

The transformation hypothesis notes that sufficiently intimate technological integration may not simply extend human cognition but fundamentally alter the ontology of personhood [Sample et al. 2019; Kögel et al. 2019]. The transformation hypothesis emphasises that deeper AI integration may not simply enhance the human person but that the traditional understandings of identity, consciousness, and selfhood may be transformed or altered [Goering et al. 2021; Kögel et al. 2019]. While cognition may become faster or more efficient when integrated with future AI-enabled BCIs through implanted brain chips, human thought may become partially dependent on artificial systems in ways that change how memory, reasoning, and agency are experienced. This explains why Sample et al. [2019] conclude that "the development of neural technologies raises important questions about the concept of personhood and its role in society." From the perspective of the transformation hypothesis, the main questions concern the kind of person who emerges through technologically mediated cognition, rather than simply the impact of technology on the person. This is because the boundary between human thought and machine-generated cognition becomes blurred.

If the brain chip developed by companies such as Neuralink Corporation helps the human brain to receive prompts, suggestions, or memory retrieval through future AI-integrated BCIs, then the line between what originates in the person and what is supplied externally may become difficult to distinguish. In their ethical commentary on neurotechnologies, Goering et al. [2021] argue that BCI technologies raise questions about "human identity, authenticity and agency". Prospective AI-integrated BCIs may intervene in the very processes by which an individual forms judgments and intentions. This can lead to total internal dependence because an individual might end up relying on the inbuilt future AI-enabled BCIs for memory recall, deliberation, creativity and decision-making support. If this happens, then nothing can stop human cognition itself from being reconfigured around artificial mediation.

From the perspective of the transformation hypothesis, this raises an important question about human identity [Sample et al. 2019]. Is the person still the same if memory and reasoning are technologically modified? While personal identity can be affected by education, traumatic experiences, relationships, or even the acquisition of new tools, neurotechnologies are likely to alter identity in a more radical sense by intervening in the cognitive processes through which continuity of self is maintained. Yuste et al. [2017:159] argue that AI and BCIs must "respect and preserve people's privacy, identity, agency and equality." If technology can edit, extend, or outsource memory and reasoning through future GenAI-empowered implanted brain chips, then the self is no longer simply acting with tools but is actually being partially assembled through them.

Another question may be raised from the perspective of authenticity. Are AI-assisted thoughts genuinely 'ours'? The question is whether a thought is original and the person can meaningfully endorse it as their own. If future AI-enabled BCIs through implanted brain chips completes our sentences, suggests our conclusions, or anticipates our relations, then thought itself may become a collaborative event between brain and machine. According to Sample et al. [2019], neurotechnology may help or harm depending on how it is designed and governed, and therefore should be understood through "the impact of BCIs on what it means to

be a person and to be recognised as such by others." The argument is crucial because authenticity is not only inwardly felt but also must be socially recognised. If someone whose cognition has been enhanced through future AI-enabled BCIs such as brain chips is perceived as less autonomous, accountable, or authentic, then the social meaning of personhood changes as well.

Privacy concerns primarily sharpen these questions. UNESCO [2025] states that "neurotechnology also raises serious ethical concerns - especially as it can access and manipulate brain activity, revealing personal information about identity, emotions, and thoughts. When combined with artificial intelligence, it poses risks to human dignity, autonomy, and mental privacy." BCIs might lead to the hacking of a human brain, which, apart from being an ethical concern, remains a genuine cybersecurity and medical risk. Because future AI-integrated BCIs are likely to rely heavily on wireless communication protocols, they will remain vulnerable to the same exploits as any other connected Internet of Things (IoT) device.

Let us imagine a scenario where a hacker takes complete control of an individual's mind and reads his or her private thoughts through a brain chip. This is linked to the notion of brainjacking, which "refers to the exercise of unauthorised control of another's electronic brain implant" [Pugh et al. 2018:219]. The ethical concern of intercepting thoughts or changing therapeutic brain stimulation is crucial because neurotechnology has already moved closer to the formation of thought itself [Pugh et al. 2018; Garza-Herrera 2024; Mihić Zidar et al. 2026]. If that happens, then the inner self becomes technologically decipherable, vulnerable, and potentially exposed. In that case, personhood is transformed by the possibility of external access to mental life. For this reason, the main question the transformation hypothesis asks is whether integrating artificial intelligence through BCIs into cognition creates a new form of technologically mediated personhood.

Neurotechnologies such as AI-integrated brain enhancement do not simply sit beside the person as neutral tools, but actually enter the domain of memory, intention, identity, and self-recognition. Sample et al. [2019] argue that "the hopes and fears associated with BCIs can be productively understood in terms of personhood." Rather than simply enhancing the human person, future AI-enabled BCIs such as implanted brain chips may transform conventional understandings of identity, consciousness, and selfhood. The foregoing discussion does not mean that the transformation hypothesis is necessarily anti-technology. Rather, it emphasises that enhancement redefines what counts as human agency, what is experienced as authentic thought, and how the self is imagined in relation to machines [Yuste et al. 2017; Pugh et al. 2018]. Perceiving AI-integrated cognition as normal also involves understanding the human person as hybrid and technologically co-constituted rather than being biologically 'purely human'. This also means interrogating whether the transformed self can still be recognised as a dignified, responsible, and free person.

3. Human Dignity as the Ethical Measure

The extended mind and transformation hypotheses help to understand what neurotechnologies such as those being developed by Neuralink may do to human capacities and self-understanding [Pugh et al. 2018; Garza-Herrera 2024]. The question now regards the major ethical standard by which those changes or transformations should be judged. Human dignity provides an appealing normative framework. This is because the central concern is whether AI-enhanced cognition remains compatible with the intrinsic worth of the human person. UNESCO [2025] states this plainly: "neurotechnology can alter the brain and mind in deep ways, so human rights and the intrinsic value of each person must be protected." From that perspective, we argue that technological progress should not be self-justifying but that it must be ordered toward the protection and flourishing of people. The importance of human dignity lies in the fact that it distinguishes enhancing humans as free persons from redesigning humans as optimised biological machines.

The preceding discussion demonstrates that neither the extended mind hypothesis nor the transformation hypothesis offers a sufficient ethical standard for evaluating AI-integrated brain enhancement. The extended mind hypothesis primarily explains why technological enhancement may be regarded as a continuation of human development. However, this understanding risks overlooking the deeper ethical implications of incre-

asingly human-machine integration. The transformation hypothesis rightly emphasises threats to identity and authenticity, though it somehow encourages excessive technological pessimism by treating transformation itself as ethically problematic. What is required is therefore a normative framework capable of evaluating enhancement without either celebrating or rejecting it in advance. Human dignity should be at the centre because it shifts ethical evaluation away from technological capability towards the intrinsic worth of the human person. Though technologies may reshape the expression of personhood, the intrinsic moral worth that grounds human dignity ought to remain unaltered.

One of the major concerns with AI-integrated brain enhancement regards the possibility of algorithms influencing personal decisions, thereby diluting the very basis of individual identity, autonomy and thought [UNESCO 2025; Miller et al., 2023]. With advancements in neurotechnologies, human beings should not be downgraded to the recipients of machine optimisation, as this has an impact on the very definition of human dignity. If not carefully integrated, AI-integrated enhancement may reduce human worth to measurable outputs such as intelligence, speed, memory, or efficiency. Yet, human dignity is deeper than cognitive performance. The German Ethics Council [2023] warns that "the use of AI must foster human flourishing, not diminish it…Artificial intelligence cannot substitute human intelligence, responsibility, or judgment." The caution is relevant because tech giants which promise faster learning, expanded memory, and improved decision-making can easily seduce society into treating the enhanced mind as a superior one.

Considering dignity as a measure involves not only asking whether AI makes us more efficient, but also the extent to which it actually helps us become more fully human. This is crucial because there is a difference between enhancing and redesigning a person. While enhancement focuses on supporting the person's capacity, redesigning a person as an optimised machine has ethical implications because it involves measuring one's worth based on what can be computed, accelerated, or controlled. As the Australian Catholic University [2026] states, "artificial intelligence should serve genuine human purposes rather than redefining what it means to be human." Although AI-integrated brain enhancement may increase cognitive power, if this involves subtly shifting the criteria of human worth from dignity to utility, then it risks becoming anthropologically deforming rather than morally enriching.

As an ethical measure, human dignity is directly linked to human flourishing. In their work on AI and flourishing, Barravecchio and Hajduk [2024] argue that advanced technologies are ethically defensible only "to the extent that they promote the genuine flourishing of human beings in a just fashion." In that sense, the legitimacy of BCIs, including implanted brain chips, does not primarily depend on whether they change, improve or enhance human beings. Rather, their legitimacy primarily depends on whether those changes, improvements or enhancements preserve the dignity and intrinsic value of the human person. For example, a technology that expands memory or increases productivity while weakening agency and diminishing responsibility would fail this test. AI-enhanced cognition should therefore serve human flourishing rather than redefine humanity based on technological standards [Barravecchio & Hajduk 2024; UNESCO, 2021]. If technology remains subordinate to the person while preserving privacy, freedom, moral judgment, and inner worth, then human dignity is not an obstacle to innovation but the very condition that makes innovation ethically human.

3.1 Autonomy and Agency

One of the major ethical questions posed by neurotechnologies regards whether they can preserve human autonomy when cognition becomes technologically mediated. Who thinks when human beings think with AI future AI-enabled BCIs such as implanted brain chips? Brain-computer interfaces already raise "questions about the degree of agency a BCI affords its user and the extent to which a BCI user ought to be accountable for actions undertaken via the device" [Davidoff 2020:9]. As future AI-enabled BCIs will be integrated into the human brain, it will be crucial to question the extent to which algorithms will shape preferences, choices, and beliefs at the level of thought and reasoning itself. In that sense, the boundary between assistance and influence becomes ethically decisive because human agency involves both the ability to act and the capacity to deliberate, choose, and endorse one's actions as one's own [Vlek et al. 2014; Burwell et al. 2017].

There is also a possibility of AI systems subtly steering cognition even before a decision is ever made. Catena [2026] argues that AI mediation can influence "autonomous deliberation, decision-making, and behaviour," while automation can alter "the conditions for autonomous choice and action." From the perspective of implanted BCIs, such influence may be even more severe since the technology may interact directly with neural processes, rather than merely suggesting options. In a context where one's memory recall, attention, or thought are increasingly filtered or mediated through AI, the question is whether it is the machine or the person who is actually making decisions. Although the person may still experience the decision as internal, its formation may already be partially externalised through the future GenAI-empowered implanted brain chip or another BCI technology. This challenges the traditional understanding of responsibility, which places the human person at the centre. If AI contributes, influences, or mediates a certain decision, who is morally responsible?

In the context of BCI ethics, responsibility is not a straightforward issue. Burwell et al. (2017) raise concerns about "moral and legal responsibility," including whether users have less control over thoughts than over actions and whether they are responsible for "all of the device's output." Similarly, Sun and Ye (2023) argue that "autonomy and agency are vital for individuals' ability to control their daily activities spontaneously," but that write-in BCIs may "alter users' agency and autonomy" and can even erode them if users receive signals they cannot control. Thus, if AI-generated suggestions become too authoritative in one's mind, then the user may no longer be the primary source of practical reasoning but a dependent recipient in a machine-assisted decision chain. It is crucial to ensure that such systems neither replace nor weaken the ability of human beings to make ethical and rational decisions or judgments nor hinder them from acting freely and being responsible for their own actions [Zeng et al., 2021]. The Council of Europe Parliamentary Assembly (2020) emphasised that:

> *This technology should not be used against a subject's will or in a way that prevents the subject from freely taking further decisions about their continued use. Special care will be needed where such technology is used to treat chronic pain, drug dependency or other conditions where interruption of treatment could lead to discomfort or distress… This technology should not prevent an individual from acting freely and being responsible for their actions. Human beings, acting freely according to their natural (as opposed to enhanced or symbiotic) consciousness, must remain the only decision-makers and the primary actors in society, especially in matters that may impact human rights and democratic processes.*

The formulation is important because it outlines the goal not as maximal machine contribution, but as preserved human agency, capacity, and responsibility. While a technology that helps human beings to think better is beneficial, one that gradually transfers or replaces cognitive authority from humans to machines through brain chips is unethical [Zeng et al., 2021]. Can human beings remain autonomous and responsible for their decisions when their thinking is technologically mediated through future BCIs such as implanted brain chips? The response depends on the extent to which AI integration strengthens or weakens cognitive freedom. Is the user still able to critically assess AI inputs, reject them, revise them and own the final judgment? If the answer is yes, then AI mediation may remain compatible with autonomy.

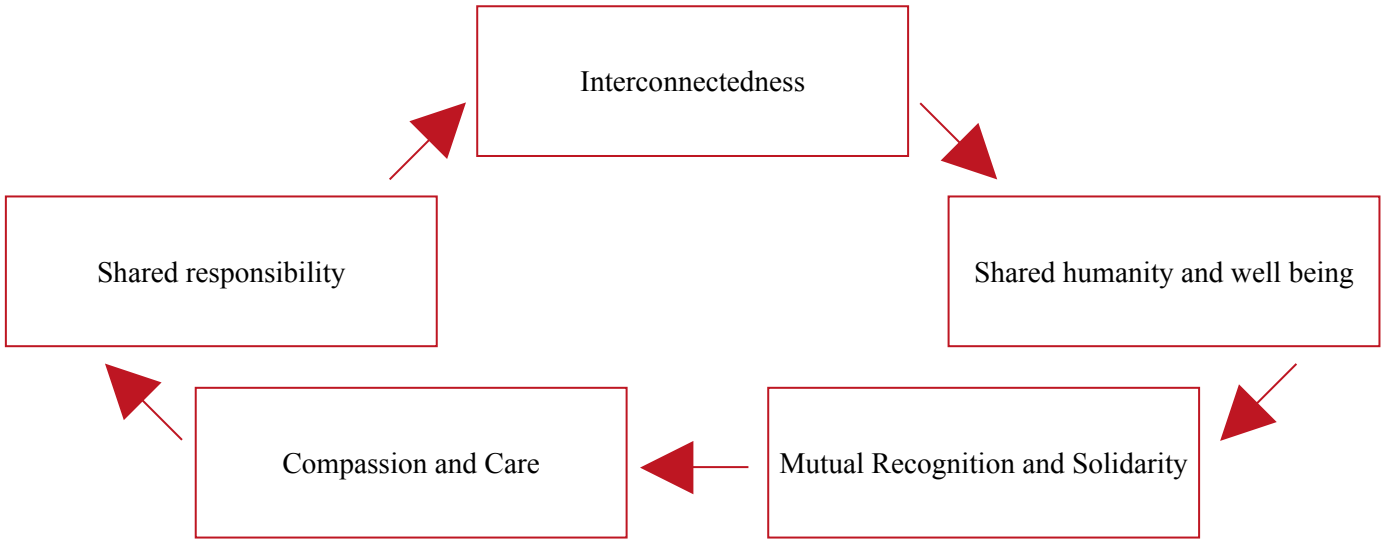
However, if the neurotechnology becomes so deeply embedded that it shapes human perception, desire, thought or belief before reflective endorsement is possible, then agency has been reconfigured [Yuste et al. 2017; Council of Europe Parliamentary Assembly 2020]. With the development of BCIs and neurotechnology in general, it is apparent that the future of human personhood depends on whether AI integration enables people to do more while remaining authors of their own lives. Once the question moves from restoration to enhancement, the challenge is not simply to regain agency, but to ensure that the enhanced mind still has agency at all. In other words, if AI integration aims at strengthening human agency, then it may enrich personhood [Gruica 2025]. However, if it slowly replaces human cognitive authority, then the very definition of human agency and autonomy can be dissolved from within. The ethical future of neurotechnology depends upon ensuring that humans remain the primary decision-makers, even when they think with machines.

4. Towards an Ethical Future of Human-AI Integration

The future of brain-computer interfaces and other implantable neural interfaces such as AI-integrated brain enhancement should not be built on either uncritical acceptance or absolute rejection. While embracing cognitive enhancement might be beneficial, the extent to which such enhancement can be governed so that it remains faithful to human dignity, autonomy, transparency, and human control remains problematic. UNESCO [2025] notes that "throughout the whole lifecycle of neurotechnology, the protection and promotion of the rights of autonomy and freedom of thought must be secured," and that neurotechnology "should never be used to exert undue influence or manipulation." Principles such as autonomy and thought are critical because they ensure that technological innovation remains answerable to the moral reality of the person.

Moreover, ethical governance must ensure that BCIs are designed and deployed in ways that preserve meaningful human oversight, informed consent, and accountability, rather than simply being a checklist of safety measures. Zeng et al. [2021:210] hold that "the autonomy of human decision making and judgment should be respected and maintained," and that such devices "should not be used to replace and weaken human decision-making ability." Through openness and transparency, tech companies such as Neuralink should ensure that risks are mitigated throughout the design, development and deployment stages [Zeng et al. 2021]. Such a process creates a conducive ethical vision for any future in which implanted brain chips remain tools for human good rather than instruments of cognitive control or manipulation.

An African philosophical perspective centred on the principle of *Ubuntu* deepens this framework in critical ways. In African thought, personhood is often understood relationally rather than as isolated self-sufficiency [Mutelo et al. 2026; Gwara & Mutelo 2025]. According to Mangena (2020), "the term *Ubuntu* emphasises the importance of a group or community, and it finds its clear expression in the Nguni/Ndebele phrase: *umuntu ngumuntu ngabantu,* which when translated to Shona means *munhu munhu muvanhu* (a person is a person through other persons)." The notion of *Ubuntu* thus emphasises the importance of shared humanity or well-being, interconnectedness, and the common good. In that sense, AI-integrated cognition enhancement should lead to human flourishing rather than being regarded as solely individual or computational. This is because the person is constituted in community, responsibility, and shared life. The diagram below shows the interconnectedness.



Ubuntu also challenges the understanding that enhancement is primarily an individual achievement. From the perspective of *Ubuntu*, the value of a person cannot be reduced to superior intelligence, enhanced memory, or increased computational efficiency. This is because personhood is fundamentally realised through relationships characterised by mutual recognition, solidarity, compassion, and shared responsibility [Mangena 2020; Gwara & Mutelo 2025]. Thus, AI-integrated cognition should be evaluated not only according to what enhanced individuals become capable of doing but also according to the extent to which such enhancement strengthens or weakens communal flourishing. For instance, neurotechnologies that isolate persons within technologically optimised individualism would stand in tension with *Ubuntu's* relational conception of humanity. *Ubuntu* therefore complements rather than replaces universal human rights by grounding dignity in relationships of mutual recognition and communal responsibility.

As tech companies such as Neuralink Corporation continue to make significant advancements, it is crucial to shift the question from what an enhanced individual can do to how enhancements are likely to affect relationships, shared purpose, solidarity, care, and the common good [Mouazen & El Bouyed 2025; Zeng et al. 2021; Mutelo 2025]. The narrow technological imagination that treats the human being as a configurable cognitive system can be challenged by such a holistic notion of personhood. The perspective shifts emphasis from reductionist perspectives that focus on speed, output, or information processing when valuing persons. The ethical legitimacy of AI-integrated cognition should be justified based on whether it leads to human flourishing communally and humanely or merely focuses on efficiency. Rather than diminishing the traditional understanding of personhood, it must strengthen the capacities by which persons relate, deliberate, and act responsibly in relation to themselves and others.

Given that cognitive enhancement may soon be widespread, the challenge lies in how human beings can embrace it without losing the qualities that make humans persons. The ethical acceptance of cognitive enhancement should depend on whether it remains subordinated to human dignity and relational flourishing [UNESCO 2025; Yuste et al. 2017]. Technologically augmenting brains does not necessarily mean that persons must be treated as isolated optimisation projects and their freedom, thought, and agency ought to be exercised based on moral relation. This requires a high level of transparency on the part of tech companies because without openness about how neural systems work, it can be challenging for users to freely consent, communities to fully trust, and to prevent moral responsibility from being blurred. As Mouazen and El Bouyed (2025) put it, "technical transparency and ethical integration are the essential conditions for responsible BCI development."

Therefore, we argue that the future of AI-integrated cognition enhancement should be guided by an understanding of humans as relational, moral, and dignified beings rather than merely enhanced cognitive systems. This calls for the need to resist any move to define human worth in terms of computational performance or efficiency alone. From that perspective, BCIs do not force individuals to choose between human dignity and embracing technological progress. The foregoing discussion suggests that future ethical governance should avoid both uncritical technological enthusiasm and reactionary resistance to innovation. Instead, responsible AI-integrated neurotechnology requires a balanced framework capable of distinguishing ethically legitimate enhancement from enhancement that compromises human dignity. Such a framework recognises that technological innovation is not ethically valuable merely because it expands human capability. Rather, its legitimacy depends upon whether it preserves authentic agency, relational responsibility, and the intrinsic worth of every person.

Conclusion

The promises of brain-computer interface technologies create an exciting but ethically demanding future. Although they have the potential to extend human capacities beyond imagination, several crucial questions about personhood, identity, autonomy, and dignity can be raised. It is beneficial that these technologies improve memory, reasoning, or creativity. However, it is questionable if they can do so without diminishing the very definition of being human. This paper has argued that AI-integrated cognition through BCIs should be judged by both technological sophistication and its capacity to serve human flourishing. Neurotechnology must aim to strengthen human freedom, responsibility, and meaningful agency, rather than replace or diminish them. However, if it diminishes or replaces human judgment, blurs moral responsibility, or reduces persons to optimised cognitive systems, then the very dignity they claim to improve is undermined. The emphasis should therefore be on ethical reflection, considerate governance, and a profound commitment to safeguarding human dignity and autonomy. Neurotechnologies should remain instruments in the service of humanity, rather than redefining or reconfiguring persons, thereby reducing them to machines. With the ongoing advancements in BCIs, the challenge is to ensure that cognitive enhancement remains humanly meaningful, morally accountable, and socially life-giving. Therefore, the ethical future of brain-computer interfaces will not be determined by the sophistication of artificial intelligence but by the moral vision that guides its development. In the face of AI-integrated cognitive enhancement, human dignity, meaningful autonomy, and relational personhood must function as the normative boundaries within which future neurotechnologies

evolve. From this perspective, the most important question is not whether humans can smoothly and safely merge or integrate with intelligent machines, but whether such integration enables persons to flourish while remaining free, morally responsible, and genuinely human.

References

Australian Catholic University. 2026. *Purpose and flourishing*. *AI Ethics Hub*. URL: https://libguides.acu.edu.au/ai-ethics/purpose-and-flourishing.

Barravecchio, M. and D. Hajduk. 2024. Virtues and vices: The ethical implications of AI on human flourishing. *Journal of Student Research* 13(3): 1–9. https://doi.org/10.47611/jsrhs.v13i3.7436.

Burwell, S., M. Sample and E. Racine. 2017. Ethical aspects of brain-computer interfaces: A scoping review. *BMC Medical Ethics* 18(1): Article 60. https://doi.org/10.1186/s12910-017-0220-y.

Catena, E. 2026. AI and human autonomy: A literature review. *AI and Ethics* 6: Article 126. https://doi.org/10.1007/s43681-025-00958-4.

Clark, A. and D.J. Chalmers. 1998. The extended mind. *Analysis* 58(1): 7–19. https://doi.org/10.1093/analys/58.1.7.

Council of Europe Parliamentary Assembly. 2020. *Resolution 2344 (2020): The brain-computer interface: New rights or new threats to fundamental freedoms*. Strasbourg: Parliamentary Assembly of the Council of Europe. URL: https://www.europeanrights.eu/public/atti/Resolution_2344_(2020)_ENG.pdf.

Davidoff, E.J. 2020. Agency and accountability: Ethical considerations for brain-computer interfaces. *Rutgers Journal of Bioethics* 11(Spring): 9–20.

Finlay, L. 2024. Let's not elevate brain tech over our humanity. *Australian Human Rights Commission*. URL: https://humanrights.gov.au/about-us/media-centre/opinion-pieces/opinion-pieces/lets-not-elevate-brain-tech-over-our-humanity.

García-Barrios, D.A. 2025. Epistemic partner or cognitive crutch: A conceptual model of cognitive offloading in human–artificial intelligence collaboration. *IEEE Internet Computing* 29(5): 7–16. https://doi.org/10.1109/MIC.2025.3626694.

Garza-Herrera, R. 2024. Humans use tools: From handcrafted tools to artificial intelligence. *Journal of Vascular Surgery: Venous and Lymphatic Disorders* 12(2): 101705. https://doi.org/10.1016/j.jvsv.2023.101705.

German Ethics Council. 2023. *Ethics Council: Artificial intelligence must not diminish human flourishing*. Press Release 02/2023, 20 March. URL: https://www.ethikrat.org/en/press/press-releases/ethics-council-artificial-intelligence-must-not-diminish-human-flourishing/.

Gilbert, S.J., A. Boldt, C. Sachdeva, C. Scarampi and P.-C. Tsai. 2023. Outsourcing memory to external tools: A review of intention offloading. *Psychonomic Bulletin & Review* 30(1): 60–76. https://doi.org/10.3758/s13423-022-02139-4.

Goering, S., T. Brown and E. Klein. 2021. Neurotechnology ethics and relational agency. *Philosophy Compass* 16(4): e12734. https://doi.org/10.1111/phc3.12734.

Gordon, E.C. and A.K. Seth. 2024. Ethical considerations for the use of brain-computer interfaces for cognitive enhancement. *PLoS Biology* 22(10): e3002899. https://doi.org/10.1371/journal.pbio.3002899.

Gruica, T. 2025. The ethical significance of brain-computer interfaces as enablers of communication. *Voices in Bioethics* 11: 89–93. https://doi.org/10.52214/vib.v11i.14149.

Gwara, J. and I. Mutelo. 2025. Reimagining children's rights in Africa: An African philosophical approach. In *Thinking Children's Rights in Africa: Emerging Issues*, edited by I. Mutelo, J. Gwara and M. Nyachowe, 17–30. Adelaide: ATF Press.

Igbalumun, A. and I. Mutelo. 2025. The African child as a person: Child rights in traditional African societies. In *Thinking Children's Rights in Africa: Emerging Issues*, edited by I. Mutelo, J. Gwara and M. Nyachowe, 3–16.

Adelaide: ATF Press. https://doi.org/10.2307/jj.35802019.7.

Kögel, J., J.R. Schmid, R.J. Jox and O. Friedrich. 2019. Using brain-computer interfaces: A scoping review of studies employing social research methods. *BMC Medical Ethics* 20(1): Article 18. https://doi.org/10.1186/s12910-019-0354-1.

Livanis, E., P. Voultsos, K. Vadikolias, P. Pantazakos and A. Tsaroucha. 2024. Understanding the ethical issues of brain-computer interfaces (BCIs): A blessing or the beginning of a dystopian future? *Cureus* 16(4): e58243. https://doi.org/10.7759/cureus.58243.

Mangena, F. 2020. Hunhu/Ubuntu in traditional Southern African thought. *Internet Encyclopedia of Philosophy*. URL: https://iep.utm.edu/hunhu-ubuntu-southern-african-thought/.

Marsh, E.J. and S. Rajaram. 2019. The digital expansion of the mind: Implications of internet usage for memory and cognition. *Journal of Applied Research in Memory and Cognition* 8(1): 1–14.

Mchia, T.T. 2023. Examining the ethics of brain-computer interfaces: Ensuring safety, the rights and dignity of personhood. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 982–991. https://doi.org/10.1145/3600211.3604733.

Mihić Zidar, L., P. Bhatia and P. Wicke. 2026. From interlocutor to cognitive extension: Brain-computer interface as a step beyond dialogue in conversational AI. In *Proceedings of the 8th ACM Conference on Conversational User Interfaces (CUI '26)*, Article 44: 1–5. https://doi.org/10.1145/3816046.3816306.

Miller, E.J., B.A. Steward, Z. Witkower, C.A.M. Sutherland, E.G. Krumhuber and A. Dawel. 2023. AIHyperrealism: Why AI faces are perceived as more real than human ones. *Psychological Science* 34(12): 1390–1403. https://doi.org/10.1177/09567976231207709.

Mouazen, B. and Z. El Bouyed. 2025. Toward ethical and transparent brain-computer interfaces: Challenges and perspectives in EEG-based AI. In *Artificial Neural Networks and Their Applications*, edited by A. Mellouk. London: IntechOpen. https://doi.org/10.5772/intechopen.1013834.

Mutelo, I. 2025. Understanding the generative artificial intelligence revolution in Zambian higher education research: Adoption, challenges, and strategies for responsible integration. *International Journal of Research and Innovation in Social Science* 9(3S): 5731–5737. https://doi.org/10.47772/IJRISS.2025.903SEDU0416.

Mutelo, I., Y. Sanyanga and T. Lemeyu. 2026. Human sexuality, cultural practice and beliefs in Africa. In *Faith, Culture and Sexuality in Africa*, edited by I. Mutelo, E. Phiri and B. Matinhira, 255–275. Toulouse: Domuni Press.

Neuralink. 2026. *Neuralink: Pioneering brain-computer interfaces*. URL: https://neuralink.com/.

Pugh, J., L. Pycroft, A. Sandberg, T. Aziz and J. Savulescu. 2018. Brainjacking in deep brain stimulation and autonomy. *Ethics and Information Technology* 20(3): 219–232. https://doi.org/10.1007/s10676-018-9466-4.

Reuters. 2023. *Musk's Neuralink to start human trials for brain implant*. Reuters. URL: https://www.reuters.com/technology/musks-neuralink-start-human-trials-brain-implant-2023-09-19/.

Sample, M., M. Aunos, S. Blain-Moraes, C. Bublitz, J.A. Chandler, T.H. Falk, O. Friedrich, D. Groetzinger, R.J. Jox, J. Koegel, D. McFarland, V. Neufield, D. Rodriguez-Arias, S. Sattler, F. Vidal, G. Wolbring, A. Wolkenstein and E. Racine. 2019. Brain-computer interfaces and personhood: Interdisciplinary deliberations on neural technology. *Journal of Neural Engineering* 16(6): 063001. https://doi.org/10.1088/1741-2552/ab39cd.

Sun, X. and B. Ye. 2023. The functional differentiation of brain–computer interfaces (BCIs) and its ethical implications. *Humanities and Social Sciences Communications* 10: Article 878. https://doi.org/10.1057/s41599-023-02419-x.

UNESCO. 2021. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO. URL: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

UNESCO. 2021. *Report of the International Bioethics Committee (IBC) of UNESCO on the Ethical Issues of Neurotechnology*. Paris: UNESCO. URL: https://unesdoc.unesco.org/ark:/48223/pf0000378724.

UNESCO. 2025. *Draft Text of the Recommendation on the Ethics of Neurotechnology*. Paris: UNESCO. URL: https://unesdoc.unesco.org/ark:/48223/pf0000393395.

UNESCO. 2025. *Ethics of Neurotechnology*. URL: https://www.unesco.org/en/ethics-neurotech.

Vlek, R., J.P. van Acken, E. Beursken, L. Roijendijk and P. Haeslager. 2014. BCI and a user's judgment of agency. In *Brain-Computer Interfaces in Their Ethical, Social, and Cultural Contexts*, edited by G. Grübler and E. Hildt, 193–202. Dordrecht: Springer.

Wander, J.D. and R.P.N. Rao. 2014. Brain-computer interfaces: A powerful tool for scientific inquiry. *Current Opinion in Neurobiology* 25: 70–75. https://doi.org/10.1016/j.conb.2013.11.013.

Yamashiro, J.K. and H.L. Roediger III. 2019. Expanding cognition: A brief consideration of technological advances over the past 4000 years. *Journal of Applied Research in Memory and Cognition* 8(1): 15–19. https://doi.org/10.1016/j.jarmac.2018.11.003.

Yuste, R., S. Goering, B. Agüera y Arcas, G. Bi, J.M. Carmena, A. Carter, J.J. Fins, P. Friesen, J. Gallant, J.E. Huggins, J. Illes, P. Kellmeyer, E. Klein, A. Marblestone, C. Mitchell, E. Parens, M. Pham, A. Rubel, N. Sadato, L.S. Sullivan, M. Teicher, D. Wasserman, A. Wexler, M. Whittaker and J. Wolpaw. 2017. Four ethical priorities for neurotechnologies and AI. *Nature* 551(7679): 159–163. https://doi.org/10.1038/551159a.

Zeng, Y., K. Sun and E. Lu. 2021. Declaration on the ethics of brain–computer interfaces and augmented intelligence. *AI and Ethics* 1(2): 209–211. https://doi.org/10.1007/s43681-020-00036-x.