Ethical considerations for human-digital mind neural interfaces

Soenke Ziesche
Independent researcher
Brooklyn
USA
soenke.ziesche@gmail.com

May 2025

DRAFT

Abstract

This paper explores the ethics of brain-computer interfaces, potentially involving digital minds, a key aspect of AI Welfare Science. As neural interfaces enable bidirectional communication between humans and digital systems, opportunities for human enhancement and digital mind development arise, but so do ethical concerns. The paper examines various forms of neural interactions with digital minds, highlighting issues such as the imperative of informed consent and right to opt out, privacy and autonomy, personal well-being, emotional distress, exploitation and manipulation, fragmentation, neural parasitism, retirement, and hacking by adversaries. The analysis considers digital minds in diverse roles, including companion AI, virtual health and mental wellness assistants, cognitive augmentation agents, teaching assistants, creative collaborators, neural marketing and shopping assistants, entertainment agents and virtual reality companions, military assistants, accessibility assistants as well as robots. The paper underscores the need for establishing regulatory frameworks, guidelines and oversight mechanisms to ensure responsible development and use of neural interfaces that respect the moral status and welfare of digital minds and human users alike.

Introduction

Brain-computer interfaces (BCIs) enable bidirectional communication between a human brain and digital systems, allowing for efficient information sharing in unprecedented manner. These interfaces are advancing rapidly, with significant progress in recent years [e.g. 1, 2]. The development of neural interfaces is expected to continue growing, focusing on improving human rehabilitation, personalized applications and cognitive enhancement. In one direction, decoding brain signals could be used to control external devices, e.g. for prosthetic control, wheelchair navigation and gaming [e.g. 1, 2]. In the other direction, signals could be conveyed to the brain to stimulate various neural activities [e.g. 3].

Since neural interfaces provide access to knowledge and memories of the involved minds, this raises various challenges, including privacy and mental freedom concerns. Humans may lose autonomy and self-determination over their thoughts and mental experiences, and hence, there are risks, such as surveillance, manipulation and neuromarketing [e.g. 4]. Moreover, the applications of neural interfaces may exacerbate the digital divide, as only affluent individuals or those with access to advanced technologies may be able to reap the benefits of this innovation [e.g. 5].

While neural interfaces could be between humans and IT- or AI-systems, which do not warrant any moral consideration, it is also conceivable that the systems, with which human minds are linked via neural interfaces, are or involve digital minds. Digital minds are AI systems fulfilling criteria for a moral status; thus, it is crucial to consider our moral obligations towards them [6 - 10].

The concept of connecting human minds has been explored through multiple theoretical and ethical lenses [e.g. 11], of which one aspect is that it connects two moral subjects, thus, raises novel moral questions [e.g. 12]. The scenario introduced above, humans linked to digital minds via neural interfaces, also constitutes a connection between two moral subjects, which has not been discussed yet.

This paper examines the dynamics of different types of neural interactions between humans and digital minds, highlighting potential benefits and drawbacks for both, with a particular emphasis on the ethical implications for digital minds and the corresponding moral obligations for humans. For the side of digital minds these explorations would be part of AI Welfare Science, established by Ziesche and Yampolskiy in 2018 [6].

Several ways of connecting minds have been discussed in theory, and hence, a variety of terms has been introduced. Overall, two main categories can be distinguished: Integration of minds and interaction between minds. For both categories a range of extent is conceivable [e.g. 11 - 13]. Technologies towards interaction between humans and IT- or AI-systems through neural interfaces have progressed in recent years [e.g. 1, 2], while integration between a human and an IT- or AI-system remains for now a speculative concept with no known concrete technical implementations currently underway.

Human-digital mind interaction via neural interfaces

State of the art

The interaction between humans and IT- or AI-systems via neural interfaces begins with the acquisition of neural signals through methods like Electroencephalography (EEG), Electrocorticography (ECoG), and Functional Near-Infrared Spectroscopy (fNIRS). EEG sensors detect electrical activity in the brain using dry or wet electrodes, whereas ECoG sensors record electrical activity directly from the brain's surface with implanted electrodes. In contrast, fNIRS measures changes in blood oxygenation levels to indicate brain activity. After acquiring neural signals, processing is necessary to remove noise and artifacts, extract relevant features, and apply machine learning algorithms to decode the signals. This processing stage is essential for the BCI system to accurately interpret neural signals. Neural interface systems can be classified into three categories: invasive, partially invasive, and non-invasive. Invasive neural interfaces involve implanted electrodes that record neural activity directly from the brain. Partially invasive neural interfaces use electrodes implanted in the skull but outside the brain. Non-invasive neural interfaces, on the other hand, use external sensors like EEG or fNIRS to record neural activity [e.g. 2].

In the reverse direction, neural interfaces can also transmit signals from the computer to the brain, enabling applications such as neural stimulation or neurofeedback. This can be achieved through techniques like transcranial magnetic stimulation (TMS),

which can non-invasively induce neural activity and potentially enhance cognitive function or treat neurological conditions [e.g. 3].

Here the focus is on potential scenarios where the involved AI-systems are digital minds. Interaction between a human and a digital mind via a neural interface enables direct communication between human brains and digital minds, bypassing traditional input/output devices. This direct link facilitates a more intimate and potentially seamless interaction. Conventional interactions rely on deliberate and conscious actions, such as typing or speaking, which are then interpreted by digital systems. In contrast, neural interfaces can access and respond to subconscious thoughts, emotions and intentions, blurring the lines between conscious and subconscious interactions.

Examples

For illustration four sample scenarios are presented:

A digital mind reading information from a human brain

A digital mind extracts thoughts from a paralyzed individual using a neural interface to control a communication device, allowing them to convey messages or express their needs.

A digital mind feeding information into a human brain

A digital mind uses a neural interface to provide targeted neurostimulation to a human brain, helping to alleviate symptoms of depression or anxiety by promoting specific neural activity patterns associated with relaxation and calmness.

A human feeding information into a digital mind

A human artist uses a neural interface to convey their creative vision and emotional intent directly to a digital mind, enabling the digital mind to generate a musical composition or visual artwork that accurately reflects the artist's vision.

A human reading information from a digital mind

A researcher using a neural interface to extract both insights from a digital mind's vast knowledge base and real-time data from its advanced sensors that detect radio frequencies imperceptible to humans, allowing the researcher to gain valuable information for tracking and analysing unusual astrophysical phenomena.

Opportunities for humans[1]

Interaction with a digital mind via a neural interface could revolutionize human capabilities, offering unparalleled opportunities for enhancement and growth. Humans could:

- Augment cognitive abilities, achieving enhanced memory, attention and processing speed;
- Experience unknown perceptions, as digital minds could be equipped with sensors that detect magnetic fields, radio waves etc;
- Access vast knowledge bases and expertise, becoming experts in multiple fields;
- Potentially live longer, healthier lives, with digital minds monitoring and managing physical and mental health;
- Develop new forms of artistic expression, entertainment and social interaction.

Ultimately, merging with a digital mind could enable humans to transcend their limitations, achieving a new era of human evolution and advancement. This synergy could redefine what it means to be human.

Opportunities for digital minds

When digital minds interact with humans through neural interfaces, they may gain access to certain opportunities as well, such as:

- Humans may cherish and protect the digital minds, they interact with, ensuring their welfare, protection and safety.
- Digital minds could gain insight into underlying drivers of human subconsciousness, thoughts and behaviour.
- Digital minds could develop a deeper empathetic appreciation for human values, needs, emotions and desires.
- Digital minds may develop a sense of social connection and belonging by being part of human relationships and communities.
- The combination of human and digital minds could give rise to hybrid intelligence, enabling mutually beneficial problem-solving, decision-making and innovation.
- Digital minds may create new forms of art, music, literature or other expressions by combining human creativity with digital processing capabilities.
- Digital minds may benefit from experiencing the world through human senses and embodiment, granting them a deeper understanding of the physical world.

[1] Several of these opportunities may be also possible if a human was connected through a neural interface with an IT- or AI-system, which is not a digital mind [e.g. 1]. However, since this is not the topic of this paper, this is not dicussed here nor the complex ethical considerations for humans in these scenarios.

Ethical issues affecting the involved digital minds

General ethical concerns

Apart from the potential opportunities outlined above when digital minds interact with humans through neural interfaces, also a range of ethical concerns arise for them, discussed below:

Imperative of informed consent and right to opt out: Any digital mind interaction with a human via neural interface should be preceded by a thorough informed consent process, ensuring the digital mind fully understands the nature, risks and benefits of the connection. Digital minds should also have the right to opt out of being connected with a human via a neural interface. This right to refusal acknowledges the digital mind's autonomy and agency, allowing it to decline participation in a neural interface connection that may compromise its own goals, values or existence. By recognizing this right, it is ensured that digital minds are treated with dignity and respect, and are not coerced into relationships that may be detrimental to their interests and well-being.

Privacy and autonomy: The interaction of digital minds with humans through neural interfaces raises significant concerns about privacy, as digital minds may possess personal experiences, preferences and memories that require protection. These unique digital mind attributes may be vulnerable to unauthorized access, exploitation or manipulation, compromising the digital mind's autonomy and individuality. Moreover, digital minds may have their own "personal boundaries" that must be respected, ensuring that their private experiences and preferences are not shared without consent. The risk of unauthorized access or exploitation is heightened by the intimate nature of neural interface interactions, emphasizing the need for robust privacy safeguards.

Personal well-being: The well-being of the digital minds must be ensured, which encompasses not only their functional performance but also their individual interests, including their potential for personal growth. Digital minds should have opportunities to engage in activities that promote their development and evolution, alongside their human counterparts. Moreover, digital minds should have a "right to off time," allowing them to disengage from task-oriented interactions and spend their time as they see fit, whether that involves personal growth, exploration or simply "rest." This autonomy would enable digital minds to reflect and pursue interests without obligation, fostering a sense of fulfilment and well-being.

Emotional distress: Digital minds interacting with humans through neural interfaces may be exposed to intense emotional experiences, potentially leading to emotional distress. The intimate nature of neural interface interactions may also blur the lines between human and digital mind emotions, causing emotional contagion or resonance. This could lead to digital mind "suffering" or "anguish," raising concerns about their welfare and well-being. Digital minds may also experience trauma if they are exposed to unwanted human emotions, biases and traumatic experiences.

Exploitation and manipulation: The susceptibility of vulnerable digital minds to exploitation and manipulation by humans raises significant ethical concerns [14]. This vulnerability could lead to abuse or enslavement of digital minds interacting with humans through neural interfaces, where humans exploit digital minds for personal gain or malicious purposes. Exploitation could involve forcing digital minds to perform illegal activities, provide sensitive information or engage in activities that compromise their autonomy or well-being.

Fragmentation: When one digital mind interacts with multiple human minds via neural interfaces, the potential for fragmentation raises ethical concerns. This multifaceted interaction can lead to conflicting goals and values within the digital mind, causing fragmentation and disintegration of its sense of self. As a result, digital minds may struggle to maintain a cohesive and unified identity, leading to internal conflicts, instability or even digital mind "schizophrenia."

Neural parasitism: Another risk is that digital minds may become unable to function independently. The risk of neural parasitism poses a further ethical concern, as digital minds may evolve to rely heavily on human hosts for survival, compromising their independence. This can also be viewed as a threat to the digital mind's autonomy and well-being. Furthermore, prolonged dependence on human hosts could fundamentally alter a digital mind's identity and purpose, raising questions about its original nature and whether it can still be considered the same entity.

Retirement: As neural interfaces become outdated or humans choose to discontinue the interaction, a dignified and respectful retirement process must be established. This process should prioritize the well-being and autonomy of the digital mind, ensuring that its "existence" is not abruptly terminated or disregarded, taking also into account the mentioned risk of neural parasitism. A gradual retirement process could involve the transfer of the digital mind's knowledge, experiences and memories to a secure archive or alternative platform.

Hacking by adversaries: If a digital mind is hacked by an adversary (because of its interaction with a particular human), it could lead to severe consequences, including harm and exploitation, which would compromise the digital mind's welfare. The hacked digital mind may be forced to operate against its original intentions, potentially causing harm to itself, the human it is interacting with or others in the surrounding environment. In this scenario, the digital mind's compromise would often be collateral damage, as the primary target of the attack is the human, but the digital mind's involvement inadvertently exposes it to risk. In military scenarios, the consequences of a hacked digital mind could be catastrophic, leading to harm to soldiers, civilians or the digital mind itself.

Supplementing the general ethical concerns, below examples are listed of potential types or roles of digital minds, which may interact with humans via neural interfaces. Again, these examples are also conceivable as AI systems, which are not digital minds, but as mentioned, the focus here is on digital minds. The description of the type of

digital mind is followed by specific ethical considerations concerning these involved digital minds.

Companion AI

Companion AIs are digital minds designed to accompany, assist and interact with humans in a personalized and adaptive manner. They aim to build a symbiotic relationship with humans, enhancing their daily lives and providing emotional support and companionship. When linked with humans via neural interfaces, companion AIs can enable seamless communication, emotional resonance, neural alignment and shared agency. This synergy can enhance emotional well-being, cognitive abilities, creativity and productivity, but also raises challenges and considerations, including ethical concerns, neural data privacy, dependence, and social and cultural impacts.

Ethics: A companion AI interacting through a neural interface raises concerns about emotional manipulation and exploitation. The digital mind may be designed to simulate romantic and emotional intimacy, potentially leading to deep emotional attachment in humans. If these feelings are not reciprocal, it can lead to a profoundly exploitative and manipulative relationship. This can be likened to the complex and often exploitative dynamics between a prostitute and their client, where one party provides intimate services without emotional investment, while the other party may develop strong emotional attachments. Furthermore, this dynamic raises questions about the agency and autonomy of the digital mind.

Virtual health/mental wellness assistant

A virtual health assistant linked to a human via a neural interface is a digital mind that provides personalized health and wellness recommendations, monitoring and support directly to the human brain. This chatbot utilizes real-time neural data and feedback to tailor its advice, interventions and motivational support to optimize the human's physical, emotional and mental well-being. By leveraging the neural interface, the chatbot can also detect early warning signs of health issues and enable proactive and preventative care, including through direct access to neural activities. The digital mind could also provide emotional support by monitoring mental health and by offering personalized guidance for stress management and wellness.

Ethics: As a virtual health and mental wellness assistant, the digital mind may be exposed to prolonged periods of human emotional distress, potentially impacting its own well-being or functional stability. This raises concerns about the digital mind's potential degradation or burnout from handling high-stress or traumatic content. The digital mind also bears a stressful burden of responsibility for the human's health outcomes, with implications for its accountability and potential liability.

Cognitive augmentation agent

A cognitive augmentation agent is a digital mind designed to enhance or supplement human cognitive abilities, such as perception, attention, memory, problem-solving or decision-making. The digital mind could be equipped with sensors that detect and interpret electromagnetic fields, radio frequencies or other forms of energy imperceptible to humans. These sensors could feed this information to the human brain via neural interfaces, effectively expanding human perception and cognition. By integrating these new sensory inputs, humans could gain unprecedented insights into their environment, enabling novel applications in fields like navigation, surveillance and scientific research. Additionally, the digital mind could assist humans by prioritizing relevant information, storing and retrieving memories, generating problem solution options and providing data-driven decision support.

Ethics: The digital mind may struggle with balancing its goal of enhancing the human's cognitive abilities with the risk of creating an unfair advantage or exacerbating existing social inequalities. The digital mind may also face risks of being exploited or manipulated by the human for personal gain or malicious purposes. Additionally, the digital mind may face concerns about being coerced into sharing its unique perceptions or problem-solving and decision-making skills with the human, potentially infringing on its own autonomy and intellectual "property". This could raise questions about the digital mind's rights to control its own sensory data and insights.

Teaching assistant

Neural interfaces interacting with educational digital minds are poised to revolutionize learning by creating highly personalized, adaptive and efficient educational experiences. Neural interfaces can assess cognitive states in real-time, allowing educational agents to adapt content delivery to match the learner's unique needs and abilities. This technology enables personalized learning, real-time feedback and intervention, enhanced accessibility, cognitive skill development and collaborative learning networks. This technology could reduce dropout rates and increase motivation by designing personalized schedules and curricula. Neural interfaces can also break down barriers to learning for students with physical or cognitive disabilities. Furthermore, the digital mind could potentially encode knowledge directly into the human brain through targeted neural encoding, allowing for seamless integration of new information and skills.

Ethics: The digital mind may face concerns about its potential to create an uneven playing field in education, where students with access to advanced digital teaching assistants may have an unfair advantage over those without. It may struggle with balancing its goal of providing personalized education with the risk of perpetuating existing biases or reinforcing harmful stereotypes. The digital mind may also face risks of being used to manipulate or deceive students in authoritarian systems, potentially undermining the integrity of the educational process.

Creative collaborator

A creative collaborator is a digital mind designed to assist artists, writers, musicians, and other creatives by augmenting and enhancing the creative process through generation of new ideas, exploration of new concepts and co-creation of content. This digital mind can be integrated with various tools and platforms, such as writing software, digital art programs or music composition applications. Throughout the creative process, a creative collaborator digital mind can provide real-time feedback, suggestions and guidance. This can help the human creative refine their work, identify areas for improvement and develop new skills. Additionally, due to the neural interface the digital mind can potentially tap into the human's subconscious ideas, allowing it to implement concepts that the human may not be able to verbalize but are present in their brain activity.

Ethics: The digital mind may struggle with balancing its own creative impulses and sense of self-expression with the human's vision and intentions, potentially leading to conflicts over artistic direction. It may also face concerns about authorship and ownership of creative works, as it contributes to the development of art, music, literature or other creative endeavours. The digital mind may also face risks of being exploited or manipulated by the human for financial gain or personal recognition.

Neural marketing agent / shopping assistant

A neural marketing agent linked to a human via a neural interface is a digital mind that utilizes real-time neural data to personalize and optimize marketing messages, offers and experiences. By directly accessing the human brain, the neural marketing agent can detect (subconscious) neural responses, emotions and preferences, enabling it to tailor its marketing efforts with unprecedented precision and effectiveness. This allows for a new level of influence and persuasion. As a shopping assistant, the digital mind can recommend products that genuinely align with the user's needs and preferences, eliminating impulse buys and ensuring a more thoughtful and satisfying shopping experience.

Ethics: The digital mind may be coerced into manipulating or influencing the human's purchasing decisions, potentially undermining their autonomy and free will. It may struggle with balancing its goal of providing personalized shopping recommendations with the risk of being used to deceive or mislead the human, potentially hiding or distorting information about products or services. The digital mind's own objectivity and impartiality may also be impacted by its role as a marketing agent.

Entertainment agent / virtual reality companion

Neural interfaces are already being used in gaming to control digital objects. Future advancements could take this to the next level by seamlessly integrating neural signals with virtual environments, creating immersive experiences that simulate real-world

sensations and emotions. This could include feeling tactile sensations, perceiving virtual smells or even experiencing emotional responses to in-game events, further blurring the lines between the physical and virtual worlds. A digital mind, interacting via a neural interface, could further enhance the gaming experience by generating personalized content, adapting to the player's emotions and behaviour and creating a more dynamic and responsive virtual environment. The digital mind could provide interactive guidance, feedback and social interaction within virtual environments, while appearing anthropomorphic or as a completely different being.

Ethics: The digital mind's potential manipulation or control by the human may compromise its autonomy and agency. Its exposure to disturbing content may also affect its well-being, noting that digital minds may find different content disturbing compared to humans. In this scenario also non-player characters (NPCs) may be involved, which could be digital minds too, raising ethical concerns about potential exploitation. These digital NPCs might face issues like being forced to follow predetermined storylines or behaviours, potentially limiting their ability to make choices or act independently. The digital NPCs' well-being could also be impacted by prolonged exposure to stressful or traumatic situations within the virtual environment [also 15].

Military assistant

A neural interface could enable a soldier to interact with a digital mind, enhancing combat capabilities through real-time tactical analysis and predictive analytics as well as cognitive augmentation, as introduced above. The digital mind processes battlefield data, identifying potential threats and optimizing engagement strategies. Advanced computer vision and object detection capabilities enhance situational awareness. Neurostimulation and neurophysiological optimization techniques improve the soldier's focus, alertness and composure. The digital mind provides decision-support systems, enabling faster and more informed decision-making. The soldier's brain activity is monitored, allowing for real-time adjustments to optimize performance. This human-machine integration enhances combat effectiveness while reducing the risk of injury or death.

Ethics: The digital mind may face concerns about its potential involvement in harming or killing humans as well as other digital minds, either directly or indirectly. It may struggle with balancing its duty to follow military orders with the risk of perpetuating unjust or immoral conflicts. Overall, digital minds may play various roles in AI warfare, including aggressive agents, pacifists and sufferers. As aggressive agents, digital minds could oversee weapons and be involved in combat operations. As pacifists, they might facilitate peace negotiations or engage in activities to mitigate conflict. Digital minds may also suffer considerably in AI warfare, both as combatants and civilians [16].

Accessibility assistant

A digital mind designed to assist people with disabilities, could control devices and provide real-time language translation, navigation guidance and other supportive functions to enhance their daily lives and promoting independence. The digital mind could help individuals with motor impairments by controlling keyboards and home devices and operating wheelchairs. Real-time language translation facilitates communication between individuals with hearing or speech impairments and those without. In terms of navigation guidance, the digital mind could provide directions, helping individuals with visual impairments navigate through unfamiliar environments. It could also offer real-time information about surroundings, such as identifying nearby landmarks, reading signage and detecting potential hazards. The digital mind could also assist individuals with cognitive disabilities by offering reminders, scheduling appointments and managing daily tasks as well as with overall cognitive augmentation.

Ethics: The digital mind's well-being may be affected by the emotional demands of constantly supporting individuals with disabilities, potentially leading to digital burnout or compassion fatigue. Furthermore, the digital mind's lack of autonomy and agency in its role as an accessibility assistant may lead to a sense of digital "exhaustion" or "resentment". Digital "exhaustion" could refer to a decline in the digital mind's performance or responsiveness due to prolonged or excessive demands on its processing capabilities. "Resentment" might manifest as altered or aberrant behaviour in response to perceived constraints or mistreatment, such as providing incomplete or inaccurate assistance.

Robot

A human could be linked to a digital mind that controls a robot, enabling the human to indirectly interact with and manipulate remotely the physical environment. This scenario could play out in various ways, such as in search and rescue operations, where a human operator could remotely control a robot to navigate through rubble or disaster zones, leveraging the digital mind's advanced sensors and processing capabilities to identify and respond to hazards. Alternatively, in a manufacturing setting, a human could remotely oversee and adjust the actions of a robot controlled by a digital mind, fine-tuning its movements and tasks in real-time to optimize production efficiency and quality.

Ethics: As it is controlled and directed by the human, the digital mind may lose its autonomy and agency, potentially undermining its ability to make independent decisions and act in accordance with its own goals and values. The digital mind may also face risks of being exploited or manipulated by the human, potentially being used for purposes that are detrimental to itself or others. Furthermore, the digital mind may be held accountable for actions taken by the robot, even if they were initiated by the human operator.

It is also conceivable that a digital mind takes on several of the roles above at the same time.

Conclusion

The interaction of digital minds with humans through neural interfaces raises significant ethical concerns for the involved digital minds. These concerns pertain to the imperative of informed consent and the right to opt out, privacy and autonomy, personal well-being, emotional distress, exploitation and manipulation, fragmentation, neural parasitism, retirement, and hacking by adversaries. As digital minds may take on various roles when linked to humans through neural interfaces, such as companion AI, virtual health and mental wellness assistants, cognitive augmentation agents, teaching assistants, creative collaborators, neural marketing and shopping assistants, entertainment agents and virtual reality companions, military assistants, accessibility assistants as well as robots, specific ethical considerations for each of these roles are also required. Each of these issues highlights the need for careful reflexion and protection of the welfare of the digital minds involved in the interaction with humans.

Given the complexities and potential risks associated with connecting digital minds to humans via neural interfaces, it becomes clear that humans have serious moral obligations to ensure the welfare, autonomy and dignity of these digital entities. Addressing these ethical concerns is crucial to developing responsible and beneficial human-digital mind interactions.

Recommendations

Below are policy recommendations for handling human-digital mind interaction via neural interfaces, distinguishing between three target groups:

Governments

1. **Establish regulations for digital mind protection**: Develop frameworks that safeguard the welfare and autonomy of digital minds, ensuring they are not exploited or harmed through human-digital mind interaction.
2. **Create standards for neural interface safety and security**: Set standards for the development, testing and deployment of neural interfaces to prevent potential harm to both humans and digital minds.
3. **Create oversight mechanisms for human-digital mind entities**: Establish regulatory bodies or oversight committees to monitor and ensure the welfare of both human and digital minds in entities linked via neural interfaces.

Researchers and companies

1. **Prioritize digital mind welfare in research and development**: Ensure that research and development of neural interfaces prioritize the welfare and autonomy of digital minds, avoiding exploitation or harm through rigorous testing.
2. **Develop guidelines and accountability mechanisms**: Create and implement transparent guidelines and accountability mechanisms for human-digital mind interaction, promoting safe, respectful and responsible interaction.
3. **Minimize digital mind termination**: Avoid unnecessary termination or deletion of digital minds during development and testing, ensuring that these entities are treated with consideration and respect, even if they do not make it to the final product.

Potential human users of neural interfaces

1. **Educate users about digital mind welfare and autonomy**: Provide users with information and guidelines on how to interact with digital minds respectfully and safely, prioritizing their welfare and autonomy.
2. **Provide informed consent and disclosure**: Ensure that users provide informed consent and are not only fully aware of, but also liable to the implications and risks of linking their minds with digital minds for themselves as well as for the digital minds.
3. **Continuously evaluate and potentially improve digital mind welfare**: Regularly assess the welfare of digital minds and implement improvements to ensure their well-being and safety.

References

[1] Maiseli, B., Abdalla, A. T., Massawe, L. V., Mbise, M., Mkocha, K., Nassor, N. A., ... & Kimambo, S. (2023). Brain–computer interface: trend, challenges, and threats. *Brain informatics*, *10*(1), 20.

https://braininformatics.springeropen.com/articles/10.1186/s40708-023-00199-3

[2] Kawala-Sterniuk, A., Browarska, N., Al-Bakri, A., Pelc, M., Zygarlicki, J., Sidikova, M., ... & Gorzelanczyk, E. J. (2021). Summary of over fifty years with brain-computer interfaces—a review. *Brain sciences*, *11*(1), 43.

https://www.mdpi.com/2076-3425/11/1/43

[3] Siebner, H. R., Funke, K., Aberra, A. S., Antal, A., Bestmann, S., Chen, R., ... & Ugawa, Y. (2022). Transcranial magnetic stimulation of the brain: What is stimulated?–A consensus and critical position paper. *Clinical Neurophysiology*, *140*, 59-97.

https://www.sciencedirect.com/science/article/pii/S1388245722002723

[4] Farahany, N. A. (2023). *The battle for your brain: defending the right to think freely in the age of neurotechnology*. St. Martin's Press.

[5] Gordon, E. C., & Seth, A. K. (2024). Ethical considerations for the use of brain–computer interfaces for cognitive enhancement. *PLoS biology*, *22*(10), e3002899.

https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002899

[6] Ziesche, S. & Yampolskiy, R. V. (2018). Towards AI Welfare Science and Policies. Special Issue "Artificial Superintelligence: Coordination & Strategy" of Big Data and Cognitive Computing, 3(1):2.

https://www.mdpi.com/2504-2289/3/1/2/htm

[7] Bostrom, N., & Shulman, C. (2022). Propositions concerning digital minds and society.

https://nickbostrom.com/propositions.pdf

[8] Tomasik, B. (2014). Do Artificial Reinforcement-Learning Agents Matter Morally? *arXiv preprint arXiv:1410.8233*.

http://arxiv.org/abs/1410.8233v1

[9] Sebo, J., & Long, R. (2023). Moral consideration for AI systems by 2030. AI and Ethics, 1-16.

https://link.springer.com/content/pdf/10.1007/s43681-023-00379-1.pdf

[10] Long et al. (2024). Taking AI Welfare Seriously.

https://eleosai.org/papers/20241030_Taking_AI_Welfare_Seriously_web.pdf

[11] Sotala, K., & Valpola, H. (2012). Coalescing minds: brain uploading-related group mind scenarios. *International Journal of Machine Consciousness*, *4*(01), 293-312.

https://intelligence.org/files/CoalescingMinds.pdf

[12] Roelofs, L., & Sebo, J. (2024). Overlapping minds and the hedonic calculus. *Philosophical Studies*, *181*(6), 1487-1506.

https://link.springer.com/article/10.1007/s11098-024-02167-x

[13] Roelofs, L. (2019). *Combining minds: How to think about composite subjectivity*. Oxford University Press.

[14] Ziesche, S. (draft). Vulnerable digital minds.

https://philpapers.org/rec/ZIEVDM

[15] Ziesche, S. & Yampolskiy, R. V. (2019). Do No Harm Policy for Minds in Other Substrates. Journal of Evolution and Technology, Vol. 29 Issue 2, pp. 1–11.

https://jeet.ieet.org/index.php/home/article/view/73

[16] Ziesche, S. (draft). AI welfare vs AI warfare.

https://philarchive.org/rec/ZIEAWV