Data Science & Artificial Intelligence

THE BIRTH OF THE FUTURE

on Neuralink & Implanting mind chips

Raneem Haleema

Supervisor: Tessa Verhoef & Eduard Fosch Villaronga

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)

www.liacs.leidenuniv.nl

29/07/2024

Abstract

Elon Musk's Neuralink received approval from the U.S. FDA to implant its brain chip in a second person after the implant in the first paralyzed patient was discontinued. Neuralink, this advanced BCI application has been trending since the start of 2024. The company's goal is not only to help patients with disease caused by brain damage, but also to achieve symbiosis with artificial intelligence by enabling everyone to implant a chip in their brain, allowing them to communicate through thoughts.This raises different ethical concerns: from “privacy of thoughts'' to “social discrimination” to “self-identification”. Does Neuralinks chip adhere to the ethical standards on a personal, social and economical level?

To answer this question, a comprehensive literature review was conducted. The research consisted of a primary and secondary search: The primary search identified the ethical aspects related to Brain-Computer Interfaces (BCI) in general, since Neuralink falls under this field of technologies. The secondary search evaluates whether Neuralink meets these defined ethical aspects. The findings of the secondary search were quantitatively and qualitatively analyzed and discussed, highlighting the aspects that require more attention to ensure that the future use of brain-chip implants technologies (invasive BCI’s) respects human identity, privacy, and rights.

key words: Ethical aspects, BCI, Neuralink, privacy, justicy, identity, informed consent

Content

Introduction

What if you would be able to order pizza by just thinking about ordering pizza? and sending an email by just thinking about your email content. Helping a grandmother with Alzheimer to remember everything again and having a full conversation with a completely paralyzed person who can move his/her eyelids. It is no science fiction anymore. Early 2024, we woke up to the trending news: Neuralink got the FDA approval (Food and drug administration) in the USA to implement chips with 1024 electrodes in the human brain. [8] Neuralink was a company managed by professor Pedram Mohseni and Randolph Nudo. This company had the goal to find treatments for brain and spinal-cord injuries like: Alzheimer's, parkinson and paralysis. [5]

The chip promises to help patients with various forms of physical disability by connecting their brain to machines. For example the patient can control an exoskeleton with their mind or paralyzed patients could use their thoughts to control prosthetic limbs and wheelchairs. [1][5] Similarly, different studies show that BCI will allow patients with locked-in syndrome to communicate again. [1][5]
As the chip is implemented in the brain, it can restore the lost connections of neurons, this is useful for patients with degenerative disorders such as Alzheimer's. [5]

When Elon Musk, the entrepreneur known for founding Tesla and SpaceX (Elon Musk Biography, 2014), decided to join Neuralink, he gained a broader vision. His vision became to extend the connectivity of healthy minds by implementing a chip surgically into their brains, allowing them to interface directly with computers. [5] Thus not only people with neurological-mental diseases such as Alzheimer's, Parkinson and paralysis, but everyone. Elon's vision is to achieve AI symbiosis: integrating AI in the human brain will function as an extra layer to help humans make better decisions. He sees the future as a future whereby advanced AI and humans work together. [5]

There are significant concerns and controversies surrounding Neuralink's practices. Earlier this year, Neuralink faced a federal investigation for possible animal-welfare violations. Internal staff complained that the animal testing was rushed, causing unnecessary suffering and deaths. [4][15] [10] A number of monkeys got paralyzed within a week of receiving the implants, and others were left with brain inflammation and swelling. Leaked photographs and tests demonstrating the treatment of these monkeys have motivated institutes, such as the "Physicians Committee for Responsible Medicine" to call for action to stop such experiments. [12]

Despite these ethical concerns, Neuralink received the FDA approval. After implanting the chip in the brain of the first patient, the trial had to be stopped due to unexpected technical errors in the chip. Surprisingly, a second approval was granted to start a trial with a second patient! This situation raises critical questions about the ethical oversight and regulatory processes involved. [15] [10]

Elon Musk's vision for the chip also raises numerous ethical issues: who will be responsible for actions taken in this context? What about the privacy of our thoughts? Social discrimination and other ethical aspects surrounding this new AI technology.

Throughout history, the most intelligent creatures on Earth have consistently seized power from other creations. Are we now at a point where AI could potentially surpass humans, relegating them to second-class citizens? As Stephen Hawking said, "Artificial intelligence could be the worst thing that happens to humanity." We need to ensure that we always maintain control and prevent AI from controlling us. [3]

It is crucial to address the ethical challenges that arise from integrating these devices and technologies into our daily lives. [3] All these factors motivated me to conduct this research and find an answer through the literature to the following question:

"To what extent does Neuralink raise ethical concerns or possibilities for individuals and society?"

To answer the research question, the following sub-questions were defined:

1. What are Brain Computer Interfaces and how do they work?
2. What is Neuralink and how does it work?
3. What ethical aspects arise in the use of BCIs according to the literature?
4. Does research on Neuralink address these ethical aspects?

The first two questions are crucial for understanding the field of this technology and its functioning. They lay the foundation for the systematic literature search conducted and answered by the last two questions.

What are Brain Computer Interfaces and how do they work?

Neurotechnology is a technology that enables a direct connection between the brain and a technical component (electrodes, computer, etc) **[9]**. Brain Computer Interfaces (BCI) lie under this kind of technology. Neurotechnology is often used for two goals: first, recording neurological signals and translating them into technical control commands and second, applying electrical stimuli on the brain to manipulate its activity **[9]**. BCI forms a non-muscular communication pathway between the living neural tissue and the computer: The device detects the signals coming from the neurons in the brain and provides real-time feedback about the ongoing brain activity. It classifies it and sends feedback to the user **[1]**.
Measuring brain activity can be done using two different methods: non-invasive and invasive.

The non-invasive method

The electrodes detect the brain activity from the head surface without the need to be implemented in the brain. They will be placed in a cap and the patient can wear this cap and the detecting process can start (Figure 1) **[9][3]**.

This method is used for instance in patients with ALS disease: Amyotrophic Lateral Sclerosis **[9]**. At advanced stages, patients are almost completely paralyzed. The only way to communicate then is to use their eyelids. In this case, measuring the brain activity is a good choice to make communication possible. Using decoding devices the change of the activity can be measured and translated. This way, patients are able to answer YES/NO questions. Further, with practice, a computerized “typewriter” can be developed so that patients can compose sentences. This way the speech generating process finds its way from the head to the computer directly **[9]**.

Figure 1: The cap of EEG placed on the head of a patient measuring brain activity.[3]

The invasive method

In cases whereby more detailed data is required, a more detailed method is needed. For instance, in the case of controlling a prosthetic arm with an attached hand (gripper) **[9]**. This is a device with multiple degrees of freedom and requires specific detection of the neurons to operate. Implementing the electrodes in the motor cortex in the brain (the motor cortex is an area that is responsible for controlling limb movement) will be a solution in this case **[9]**. To achieve this, surgery is needed, which is why this method is called invasive.

The deeper the electrodes are implemented, the better they will detect neuronal activity and the more detailed the information about the activity will be **[17]** .
This is because the invasive method offers better spatial and temporal resolution:

- Spatial: When electrodes are placed close to or directly within the target cortical areas or subcortical structures, the BCI can detect signals from individual neurons or small populations of neurons, decode specific information, and modulate particular brain functions **[17]** .

- Temporal: An invasive implanted BCI is more resistant to electrical noise interferences or movement artifacts, and it has a higher signal-to-noise ratio (SNR)**[17]** .

What is Neuralink and how does it work?

Neuralink is a type of invasive BCI implanted in the head through surgery. The procedure is conducted by a robot developed by Neuralink **[5].** The use of a robot is necessary because the chip to be implanted consists of very small threads that are barely visible to the human eye **[5]**. The robot ensures the device is implanted in the brain without touching veins or arteries. It was tested on 19 rat models with an insertion success rate of 87.1% **[5]**.

The surgery starts with a 2 mm cut into the skull. Then, the robot will use 24-micrometer (one micrometer is one-millionth of a meter) needles to insert each thread individually, targeting specific brain zones and avoiding surface vasculature. This is done by using magnifying instruments, depth tracking, and pre-selection of the insertion sites and landmarks **[5].**

The robot can visualize the micrometer threads and the cortex of the brain accurately thanks to the use of stereoscopic cameras and image stacking from six different light modules. These are capable of illuminating at 405 nm, 525 nm, and 650 nm. This enables the robot to accurately illuminate specific locations in the brain and place the threads into the cortical surface successfully (Figure 2) **[5].**
Elon Musk announced that the operation will take about two hours, and the person will be incompletely sedated (John, 2020) **[5].**

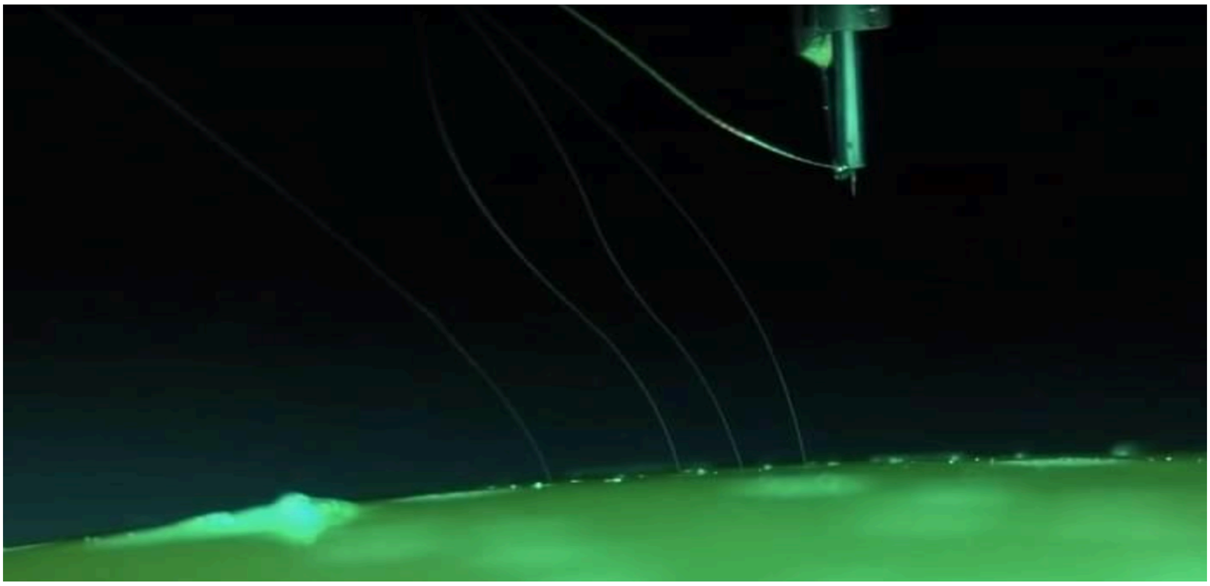

Figure 2: electrodes of Neuralink chip being inserted by the needle of the robot in the brain [5].

Electrodes

The chip detects brain signals using electrodes. It contains 96 small arrays of electrodes, each containing 32 independent arrays, summing up to 3,072 electrodes in total. Compared to older techniques using BCIs, the way the electrodes of the chip are implemented is a significant development. Older techniques, such as DBS for controlling robotic limbs and computer cursors, allowed the use of a maximum of 256 electrodes, while Neuralink today allows the implementation of 3,072 electrodes! This provides more accurate classification and interpretation of brain activity and better transfer of high volumes of data that can be read and amplified (figure 3) **[5].**

The threads of the chip are made of biocompatible polyimide over a gold, thin film trace. This allows a decreased immune response and provides a better biocompatibility **[5]**.

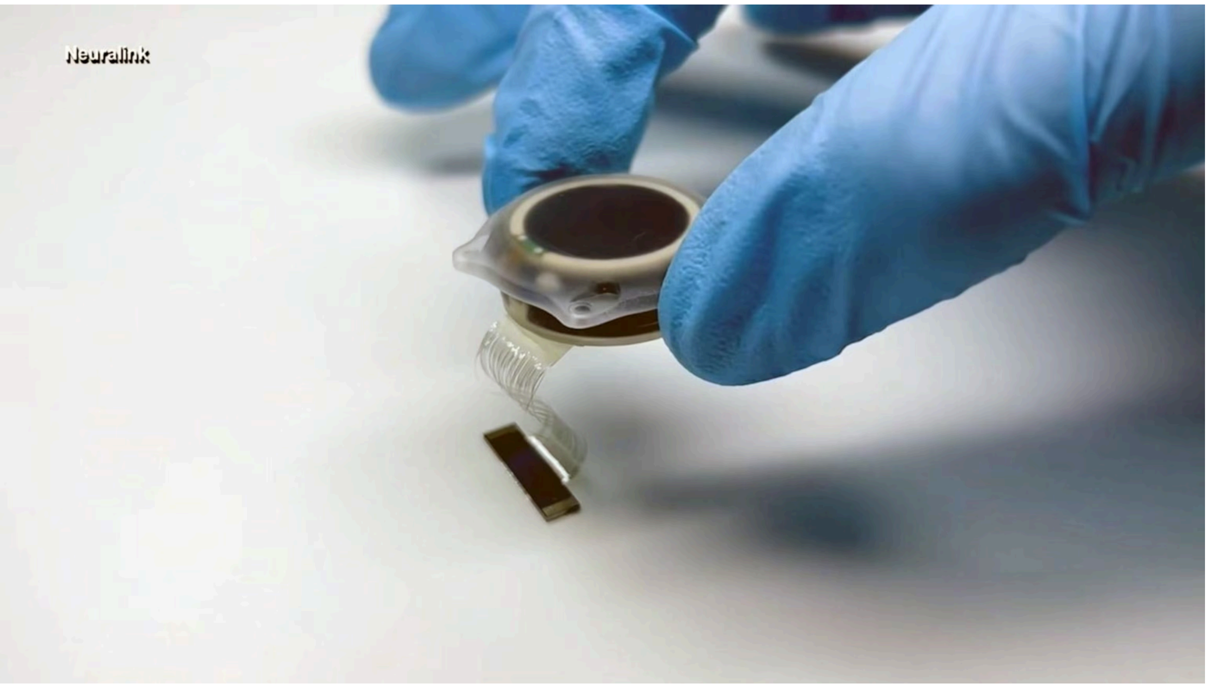

Figure 3: Neuralink chip [5].

The huge number of electrodes forms a challenge for neuralink to receive and transcribe the information. A neuron is active once it's fired and passes the signal to the next neuron. This is done by the process of neuron depolarization. This can be seen as the data from the neuron and can be algorithmically converted and stored in the Neuralink interface. This data is presented as a “spike” on the interface using a specific software and can be personally followed on the Neuralink iPhone application **[5]**.

Method

To explore the extent to which Neuralink raises ethical concerns or possibilities for individuals and society, it is necessary to first identify the ethical aspects of using brain-computer interfaces (BCIs) in general as discussed in the literature. This was accomplished by conducting a primary literature search, resulting in a list of ethical aspects that served as a foundational base. This base was then used for the secondary literature search, which aimed to compare the ethical aspects in research about Neuralink with the general BCI aspects defined by the primary search.

These two literature searches answer the third and fourth sub-questions of this research. For both the primary and secondary searches, Google Scholar was used to find relevant articles. Inclusion and exclusion criteria were applied for both searches (Figure 4).

After applying the criteria in the primary search, 25 articles were included, studied, and analyzed. The ethical aspects identified in these 25 articles formed the foundational base (refer to the document at the end of this work, following the references).

Applying the criteria in the secondary search yielded 17 articles (Figure 5). These articles were then studied and analyzed using the foundational base.

Two types of analysis were used: quantitative and qualitative. The quantitative analysis provided a clear comparison between the ethical aspects identified in the primary search and those found in the secondary search. The qualitative analysis involved discussing the identified ethical aspects within the context of Neuralink's use. This approach enabled a detailed exploration of the main research question, addressing and discussing ethical aspects while highlighting key considerations for the future application of Neuralink technology.

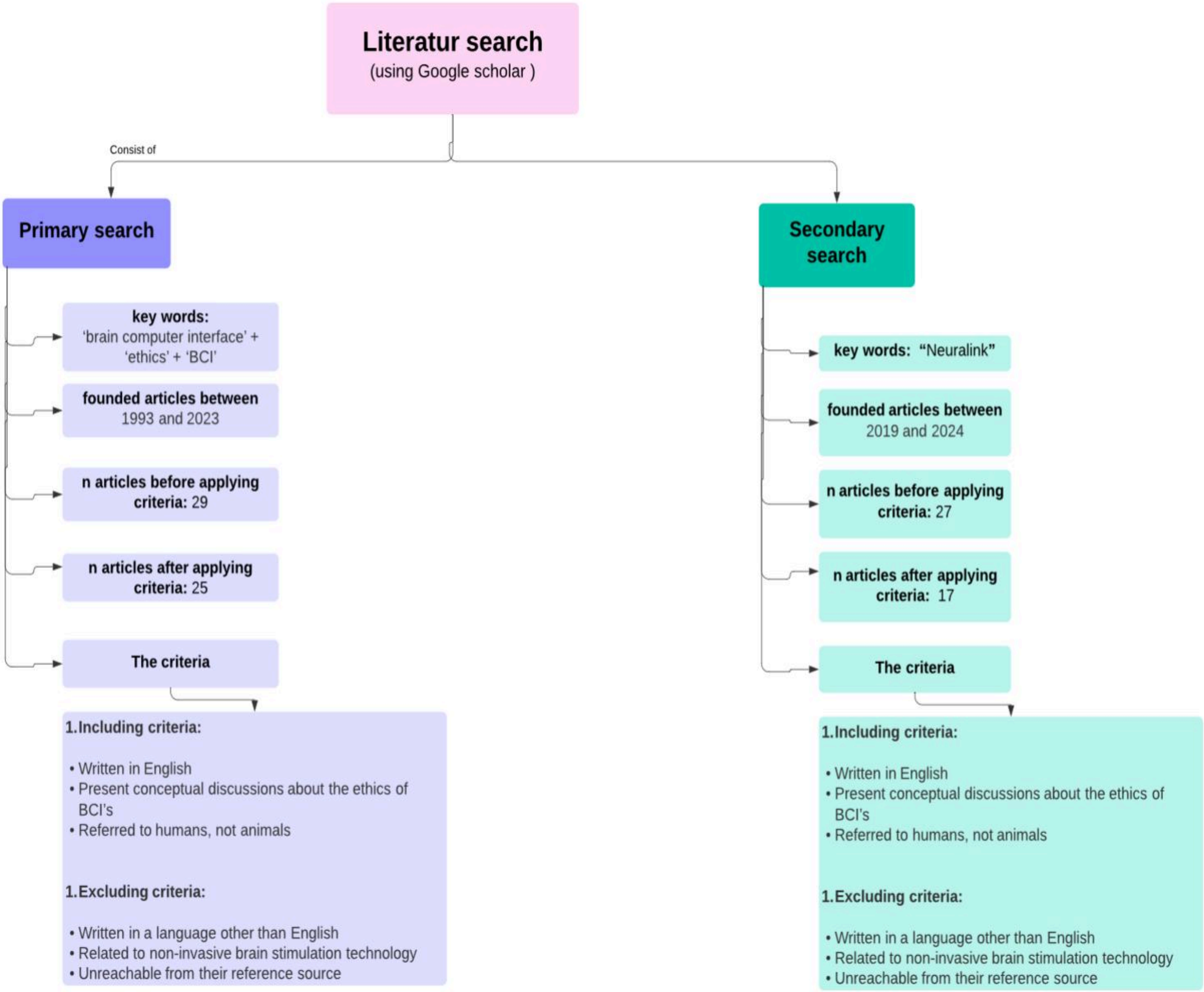

Figure 4: Explanation of the conducted literature search, including the primary and secondary search, keywords used in the search engine, and the number of articles included before and after applying the inclusion/exclusion criteria.

	Title	Reference
1	The Implications of Neuralink and Brain Machine Interface Technologies.	Armstrong, W., & Katina Michael, K. M. (2020). The implications of neuralink and brain machine interface technologies. International Symposium On Technology.
2	Elon Musk's Neuralink brain chip: what scientists think of first human trial.	Elon Musk's Neuralink brain chip: What scientists think of first human trial. (2024). Nature.
3	Why will Neuralink Change Humanity Forever?	Fadziso, T. (2020). Why will Neuralink change humanity forever? AJHAL, 7
4	Reflecting on neuralink : What is the possible role for eastern philosophy to play in emerging technologies?	Febri Utami, D. (z.d.). Reflecting on neuralink: what is the possible role for eastern philosophy to play in emerging technologies? Vienna University.
5	Remote-controlled individuals	Fermat, M. (2022). ReMote Controlled Individuals? The Future of Neuralink: Ethical Perspectives on Human-Computer Interactions. SATEPRESS.
6	Examination of Prospective Uses and Future Directions of Neuralink: The Brain-Machine Interface.	Fiani, B., Reardon, T., Ayres, B., Cline, D., & Sitto. (z.d.). An Examination of Prospective Uses and Future Directions of Neuralink: The Brain-Machine Interface. Curieus, cureus.14192. https://doi.org/10.7759/cureus.14192
7	The Hybridization of the Human with Brain Implants: The Neuralink Project	FOURNERET, É. (2020). The Hybridization of the Human with Brain Implants: The Neuralink Project. Cambridge University Press.
8	Neuralink, An ELON MUSK START UP	Kulshreshth, A., & Anand, A. (2019). Neuralink- An Elon Musk Start-up Achieve symbiosis with Artificial Intelligence. International Conference On Computing,

		Communication, And Intelligent Systems (ICCIS).
9	Elon Musk's Neuralink " An update for humans"- A conceptual review.	Lokesh, E. S. (2022). ELON MUSK'S NEURALINK "AN UPDATE FOR HUMANS" – A CONCEPTUAL REVIEW. International Research Journal Of Modernization in Engineering Technology And Science, e-ISSN: 2582-5208.
10	From Novel Technology to Novel Applications: Comment on "An Integrated Brain-Machine Interface Platform With Thousands of Channels" by Elon Musk and Neuralink	N Pisarchik, A., & Maksimenko, V. A. (2019). From Novel Technology to Novel Applications: Comment on "An Integrated Brain-Machine Interface Platform With Thousands of Channels" by Elon Musk and Neuralink. JOURNAL OF MEDICAL INTERNET RESEARCH.
11	Neuralink – Are Brain-Computer Interfaces leading us into a technological utopia?	Neuralink – Are Brain-Computer interfaces leading us into a technological utopia? (2019). Masters Of Media.
12	Neuralink: Questions arising about the universal language	NEURALINK: QUESTIONS ARISING ABOUT THE UNIVERSAL LANGUAGE. (2022). Conferences, 2.
13	Ethical Considerations of Neuralink and Brain‐Computer Interfaces	Ong, J., G.Lee, A., & Waisberg20, E. (2024). Ethical Considerations of Neuralink and Brain‐Computer Interfaces. BMES.
14	Gadamer in a weird brain	S. Lindia1, M. (2022). Gadamer in a Wired Brain: Philosophical Hermeneutics and Neuralink. Philosophy & Technology, 35:27
15	Neuralink brings light to neural care	Sitaram, S. (2022). Neuralink brings a saving light to neural care. ISSU.
16	Using Neuralink by Humans: A Process Which Brings Humanity Closer to the Future.	Using Neuralink by Humans: A Process Which Brings Humanity Closer to the Future. (2021). SPRING LINK.

Figure 5: Included studies in the secondary search

Results

Primary search results

During the literature search, papers were analyzed to identify which ethical aspects were discussed. These aspects are referred to by different names across the papers. For instance, "privacy" and "neurological privacy" address the same issue, while "responsibility" and "accountability" both refer to the party responsible for incorrect decisions. Some ethical aspects are widely discussed, while others are only briefly mentioned in a single paper. The foundational base below counts each ethical aspect as mentioned, whether it is widely discussed or briefly mentioned. This approach provides a complete view of the ethical aspects present in the field of brain-computer interfaces.

Before diving into the results of the analysis, here is a brief definition of the ethical aspects found in the literature:

Privacy and limited confidentiality

The right to control and protect personal mental and psychological information collected by BCIs from unauthorized access, misuse, or disclosure **[1][7][9][11][14]**.

Risk benefit ratio

Comparing the effectiveness of BCIs to alternative methods and considering whether using a BCI for a specific patient has more benefits than risks **[1] [13]**.

Justice

Considering social and cultural impacts using the device by ensuring fairness in the design and in providing equal access and chances **[1]**.

Autonomy

The ability to make independent choices and control one’s actions **[1] [7]**.

Identity

The impact of BCI use on an individual's self-perception and self-sense **[11] [1] [13][9] [14]**.

Dignity

Ensuring that using BCI respects and does not undermine the inherent worth and rights of individuals **[11][9]** .

Safety

Ensuring that BCI does not cause harm, infection, trauma, and long-term complications, and evaluating non-medical issues like cognitive and emotional burdens **[1]**.

Responsibility

The accountability for actions and outcomes resulting from BCI use **[1]**.

Informed Consent

Ensuring individuals fully understand the potential psychological effects of BCI use before giving a consent to use it **[11] [9][14][1]**.

Clinical Validity

Ensuring that algorithms are thoroughly tested and validated to meet safety and avoiding biases **[7]**.

Religion

Taking into account that there are religions that reject the use of invasive BCI based on concerns about dehumanization or unnatural augmentation **[7]**.

Interaction with the Media

The responsibility of scientists to ensure accurate representation of their research using social media **[11]**.

The bar chart below (Figure 6) provides an overview of the distribution of these ethical aspects, indicating how important each aspect was found in the literature. This is based on how many different studies mentioned or discussed each aspect and how many different years had studies mentioning or discussing this aspect.

On the X-axis, the ethical aspects are represented, and the Y-axis represents the ratio. The ratio is defined as follows:

$$\text{Ratio} = \frac{\text{in how many different studies is this aspect mentioned/discussed}}{\text{In how many different years was this aspect mentioned/ discussed}}$$

The higher the ratio, the more important an ethical aspect was considered by the researchers. For instance, researchers gave more attention to privacy, while aspects such as clinical validity and interaction with media were less highlighted in research on brain-computer interfaces.

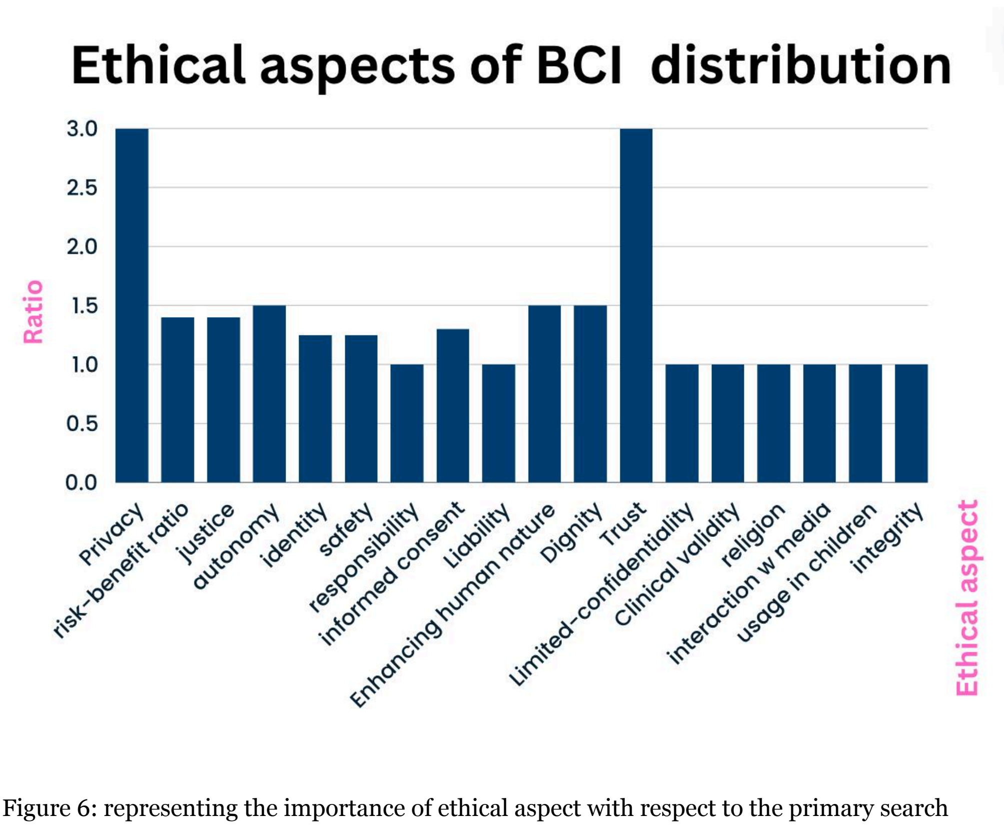

Figure 6: representing the importance of ethical aspect with respect to the primary search

Quantitative analysis

The bar chart below (figure 7) illustrates the count of each ethical aspect identified in the primary search (blue) compared to those found in the secondary search (pink).

The ethical aspects: privacy, risk-benefit ratio, justice, autonomy, identity, informed consent, trust, and limited confidentiality were less mentioned/discussed in studies on neuralink compared to studies on BCI in general.

dignity, integrity, clinical validity, and the use of the technology in children receive no attention in articles about Neuralink; these aspects are neither discussed nor mentioned.The only ethical aspect receiving more focus in Neuralink-related research, as opposed to general BCI literature, is safety.

This indicates a significant gap in the attention given to ethical issues related to the implantation of the Neuralink chip in the human brain.

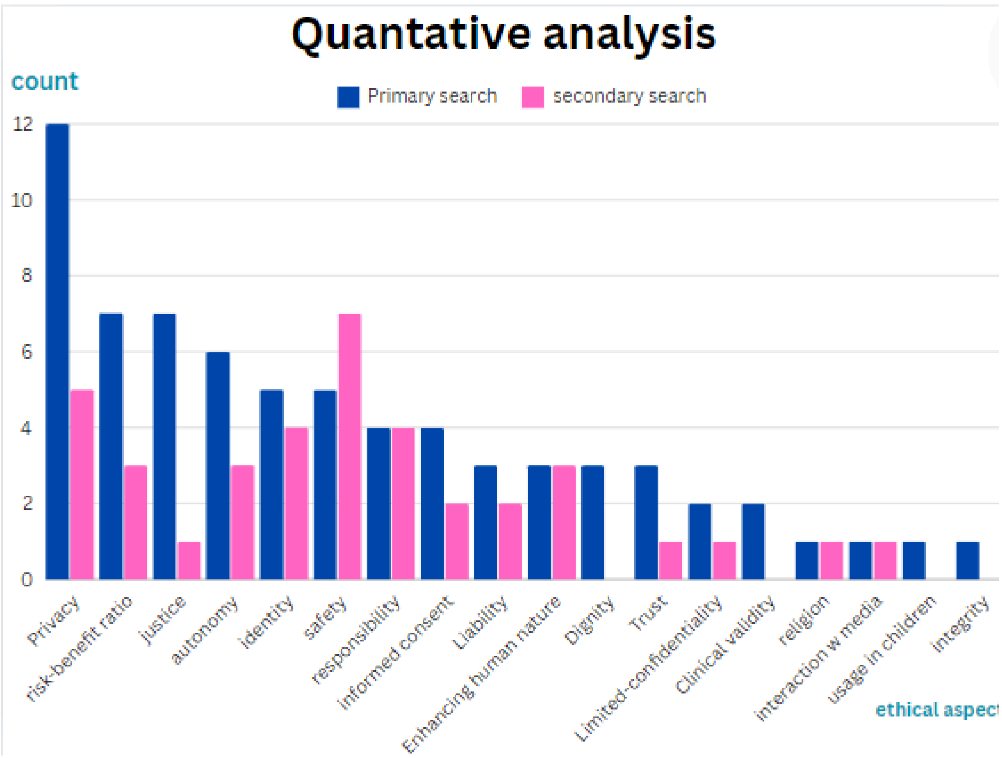

figure 7: Quantitative analysis: The x-axis represent the ethical aspects and the y-axis represent the count: in how many studies is this ethical aspect mentioned

To provide a visualization of this comparison, the table below (Figure 8) has been constructed. It represents the gaps and the extent of attention given to ethical aspects in research on Neuralink.

Aspect|| article number	1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Privacy	0	0	0	0	0	0	1	1	0	0	1	1	1	0	1	0
risk-benefit ratio	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
justice	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
autonomy	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1
identity	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0
safety	1	0	0	0	0	1	0	1	0	0	1	0	1	0	1	1
responsibility	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1
informed consent	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Liability/ reliability	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
Enhancing human nature	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
Dignity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trust	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Limited-confident iality	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Clinical validity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
religion	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
interaction w media	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
usage in children	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
integrity	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Figure 8: Vertically, the ethical aspects are represented, and horizontally, each number represents one of the studies defined in Table 5 that specifically focus on Neuralink. A specific ethical aspect is colored green (1) if it was mentioned in a study, and red (0) if it was not.

Qualitative Analysis & Discussion

privacy & limited Confidentiality & Trust

Currently, language and body language are the two moderators that help us understand and communicate with each other. BCI technology is on the way to letting us observe each other's minds more directly and decrease our privacy of thoughts. This may lead to social discrimination. For instance in the workplace or at school **[1]**.

Another point is that the chip of Neuralink is based on wireless connection. This puts it at risk of being hacked and controlled by a third party (government or unethical hackers) **[1]**. It's not only the risk of getting the chip hacked but also the risk of modifying the way it works and thus affecting the way the user will make decisions. In other words; controlling the brain of the user (Bonaci et al., 2015) **[1]**.

Beside this, the chip directly detects signals from the brain and stores them. Which and how much information is collected and stored is unknown. The type of it can be sensitive. For instance information about psychological traits and mental states such as indicating truthfulness (whether the user is telling the truth) or the user's feelings toward other people [1]. The company could have direct or indirect access to a large amount of sensitive data about our thoughts. How can we trust the company when we don't know how they might use this information—whether to change how the chip works, influence our decisions, or even control our thoughts and behavior **[self reflection]**.

From another perspective, the chip can be viewed like other media platforms (Instagram, Facebook, ..etc) or websites that store information about us. There are established protocols for handling such information, so similar detailed protocols suited to Neuralink's technology need to be established. However, the question remains whether these protocols and rules will be effectively adhered to **[self reflection]**.

This is why developing neurosecurity is crucial and needs to be given effort similar to cybersecurity and security in computer sciences (Bonaci et al., 2015) **[1]**.

Identity & Autonomy & Dignity

Directly integrating with the chip may affect the capacity for memory and self-understanding, leading the user to have questions about personal identity and sense of self. Or affect decision-making upon given information, which will lead to confusion about the ongoing psychological state and social identity **[1] [13] [9] [14]**.

The decisions made in the brain will be mostly or solely made by the chip. Allowing the chip to have this causal role will reduce the self-control of the user and thus affect the autonomy of the user (Glannon) **[1]**.

Uncontrolled use of BCIs is dangerous. It can disrupt our unwritten social rules, such as saying “please” and “thank you” in social interactions. These rules are unspoken but widely expected. (Demetriades et al., 2010) **[1]**. These social rules are the key to social cohesion and self respect. Disrupting them affects the ability to interact respectfully and lead to loss in self-dignity **[self reflection]**.

How can we find a solution for this? These ethical aspects are psychological, not practical. This means that if we want to examine the chip regarding these aspects, long-term experiments on human subjects are needed **[self reflection]**.

However, the Neuralink company did not respect the “long-term” , as one of the reasons animals were left paralyzed or with brain swelling and inflammation during the experimental phase of the chip was because Elon Musk wanted to expedite the process **[4][15] [10]**.

Even if the company takes the long term into account, the experiment itself is an ethical dilemma: how can we justify risking the psychological well-being of humans for an experiment? **[self reflection]**.

Responsibility & Trust

As Musk claimed the chip would help us make better decisions, the question arises: when taking a certain decision, who is responsible for the action taken based on this decision? The user or the developer?
In crime scenarios, a crime committed using a chip in the head makes it difficult to determine whether the criminal had the intention to commit the crime or if the action was based on a decision given by the chip. How can we control that?[self reflection] To answer this question, we need to consider "Trust." Can we use the information stored in the chip as a resource to determine the criminal's intention?

Again, to answer this last question, we need to define "intention" first. What is the intention in terms of detected neurological signals? Despite the amount of knowledge we have about the brain, researchers in psychology and cognitive neuroscience still define the brain as a black box: "If I had it all to do again, I would still call the mind a black box" (Dr. Skinner, 1987) **[9]**. "I would not use any of the new techniques for measuring information processing and the like. My point has always been that psychology should not look at the nervous system or so-called mind—just at behavior." (Dr. Skinner, 1987) **[9]**.

If we can't measure intention in terms of neurological signals because our brain is very complex and still defined as a black box in cognitive science, we can't trust the chip and the information stored in it to determine who is responsible for a decision made and a criminal action taken **[self reflection]**.

A number of ethicists and experts in moral philosophy view Neuralink as any other technology and, regarding this viewpoint, they suggest some points on how to proceed with this technology when discussing responsibility **[1]**:

- User Responsibility: The user is responsible for every action, similar to driving a car. Driving can be very dangerous, and you need to learn how to use a car properly. The same applies to the implanted chip. A majority of people believe that users of BCIs (Brain-Computer Interfaces) should be responsible for the actions performed with the aid of the chip. This was the result of a survey conducted by researchers (Nijboer et al) **[1]**.
- Manufacturer Responsibility: In the case of an accident, it could be considered a defect of the device, and the producers should be responsible for the unintended actions. This can be compared to the responsibility of manufacturers for defective products **[1]**.

In contrast, some ethicists and experts in moral philosophy argue that Neuralink is not just an advanced technology that can be treated like any other. This kind of BCI has unique features that differentiate it from other technologies, making it difficult to assign responsibility for wrong actions (O'Brolchain and Gordijn, 2014)**[1]**.
For instance, BCIs read signals directly from the central nervous system, bypassing usual body movements. As a result, actions could happen simply due to subconscious events or passing thoughts **[1]**.

JUSTICE & Enhancing human life

Neuralink is still in the clinical research stage, which overlaps some justice concerns with research ethics: Once a study is completed, what will happen to the subjects? Are they allowed to keep the chip implanted? If so, who is going to maintain the device? **[1]**. Taking away the device after completing the experiment on the subject can have an emotional impact; giving the patient hope and then taking it away may cause depression in patients, and this is an important point to address **[1]** . This situation indeed happened to the first human subject, Noland Arbaugh, when the company needed to extract the chip from his brain after implanting it due to threads detaching. Some of the threads were removed from their location and were unable to detect signals as they were supposed to do **[6][16]**.

The chip can enhance the health status of the patient. But once it becomes available to patients and non-patients, and succeeds in enhancing the user's life by, for instance, allowing them to speak a language without needing to learn it (the data of the language will be uploaded into the chip implanted in the user's brain), the question becomes whether everyone will have fair access to the chip. It may be too expensive for people with unstable economic situations, which leads to social inequality and creates social stratification **[1][7]** .

For those who decide to implant the chip and take advantage of its benefits, the question is: how long will these advantages last? The invasive methodology is risky and likely can only be done once by the individual (Klein, 2016) **[1][7]**. Does the chip need to be replaced every few years, for example? This requires surgery, and how much will this surgery cost?**[self reflection]**.

This is why reflecting on the decision to proceed with the implantation is important. Herein lies the importance of well-informed consent **[self reflection]**.

Last point to discuss is that the life experience of the user depends on the decisions taken throughout the years. These decisions depend on different factors including beliefs and religion, place of growing up, traditions, personality, and prior life experiences. These factors contribute to the uniqueness of individual choices and perspectives. How can a device take a role in that? a device programmed by an external entity determines what a better decision to make and action to take? knowing that the concept of a "better" decision is subjective and varies from a person to another based on their values and the factors called **[self reflection]**.

Informed consent & Risk-benefit ratio & Interaction with media

The risk-benefit ratio is comparing whether the implantation of the chip will have more benefits than risks for the patient. For instance, locked-in patients are non-communicative and have impaired capacity to consent (Klein,2016) **[1][13]** .

Even if the technology is beneficial and the locked-in patient agrees to have it implanted in their brain, we cannot always be sure if they truly mean what the speech generator conveys: They are unable to speak directly, which may result in the chip misinterpreting their feelings and transmitting these signals to the speech generator on the computer. For example, the system might output “I prefer to be alone” as “I want to communicate.” This misinterpretation could unintentionally worsen their feelings of being locked in **[self reflection]**.

Overall, the majority of researchers find that, in the case of locked-in state patients, the benefits of using non-invasive BCIs outweigh the risks **[1] [13]**.

For fully paralyzed patients or people with difficulties in decision making because of a damage in their brain, it becomes challenging to let them make a decision regarding the implant of the chip (Farisco et al, 2015) **[1]**. To solve that, researchers defined three key aspects to be included by informed consent:

1. disclosure: The patient has understood all the needed information **[1]**.
2. Capacity: The ability of the patient to make a reasonable decision based on the understood information **[1]**.
3. Voluntariness: The decision of the patient is made freely without any influence **[1]**.

BCI technology might not work for 15-30% of individuals, and patients may give consent because they mistakenly believe that BCI will cure them. This can lead to depression if their high expectations are not met **[13]**.

The media can contribute to this therapeutic misconception. Journalistic channels and social media are tools for researchers to communicate with the public. However, due to inaccurate or unclear explanations and communication through social media, the technology is often described as a "cure" or as capable of "reading minds," which significantly overstates its capabilities. This can lead to misunderstandings and result in consent that is not well informed **[11]**. Therefore, more attention needs to be given to the way this technology is represented on social media to ensure accurate and clear communication **[self reflection]**.

Clinical Validity, Safety

The only ethical aspect receiving more focus in Neuralink-related research, compared to general BCI literature, is safety. The chip is classified as a third-class medical device, which requires the highest level of protection and security because it is implanted in a very sensitive and complex area: the brain. This security is necessary to prevent risks such as excessive bleeding during the surgery **[5]**.

Elon Musk stated that the ultimate goal of Neuralink is to achieve AI symbiosis, allowing AI to work together with the human brain, rather than just enabling paralyzed patients to control their devices. It is important to note that neural networks in machine learning algorithms are inspired by the workings of the brain. These algorithms require thorough testing and validation. Deep neural networks (DNNs), which are used for these algorithms, are often described as “black boxes,” making it challenging to inspect them before and after use. This affects accountability in case something goes wrong **[7]**.

Another point is that AI models can introduce uncertainty instead of clarity. Given the complexity of the brain, allowing AI models to identify patterns without a specific hypothesis may lead to data misinterpretation **[7]** . Causation can be confused with correlation: is one pattern caused by another, or are they simply occurring simultaneously? **[7]** . This raises concerns about the process of detecting neurological signals and the safety of using the chip (LeCun et al., 2015; Vieira et al., 2017) **[7]** .

A further challenge is that using AI models and algorithms to detect brain activity can lead to overfitting, which reduces the generalizability of the model due to biases during training (Dietterich, 1995). Thus, even when AI models perform better than humans, they can still make mistakes (Nguyen et al., 2015) **[7]**.

In clinical applications or when testing clinical hypotheses, neuroscientists are expected to carefully evaluate the output of AI models and consider the abstraction of the model to avoid making unsupported claims about human brain function

(Hassabis et al., 2017) **[7][5]**. However, when making the chip available for non-patient also, who will evaluate the output? This raises further concerns about trusting the chip and the safety of the device **[self-reflection]**.

The effectiveness of the chip depends on the patient's condition. As explained in the introduction, BCIs have been used in paralyzed patients, whose paralysis results from damage to the brain's motor cortex. When this damage is severe, the neurological signals from this part of the brain may be very weak or unclear, making it difficult for a BCI to detect these signals. Thus, the effectiveness of the chip depends on the patient's health state. This raises a crucial question: to what extent is this new technology efficient? Is it truly worth all the effort and energy invested? More research and development are needed in the future **[5]**.

Conclusion

Although Neuralink offers hope for paralyzed patients, it is not suitable for all cases, particularly for those in locked-in states. Researchers agree that non-invasive BCIs are generally better for locked-in patients compared to invasive ones like Neuralink.

The company's goal is to achieve AI symbiosis, where the chip becomes an integral part of daily life and decision-making. The literature review revealed a significant gap in addressing the ethical aspects that are crucial for the use of the chip. These aspects, which are fundamental in the field of invasive BCIs, must be considered before making the chip available for widespread implantation to achieve the company's goal. Starting from privacy, since the chip detects sensitive data from the brain and relies on wireless connections, making it vulnerable to hacking and misuse. To solve this, strong security protocols and clear regulations must be put in place.

Personal Identity and Dignity are also at risk. The chip could impact how users perceive themselves and their dignity. Long-term studies are required to understand these effects better, but such studies pose ethical dilemmas as they risk the psychological well-being of participants.

The chip's involvement in decision-making processes could affect users' self-control and reduce self-autonomy. Detailed studies on how the chip will influence decision-making are needed before it can be widely approved.

Further, using the chip, it becomes a challenge determining who is accountable for actions taken using the chip. It is essential to develop clear rules and possibly new laws to address responsibility and ensure fair judgments.

If the technology proves effective, there are concerns about its impact on social inequality. Not everyone may be able to afford the chip, which could lead to discrimination in workplaces and schools. Addressing this issue is crucial to prevent social division and maintain justice in the community.

Finally, the engineering and validation of AI models used in the chip must be trustworthy. Machine learning engineers need to ensure that algorithms are reliable given the complexity of the human brain.

In conclusion, the unresolved questions from this research must be answered before granting approval for the chip's use by patients and non-patients. This calls for collaboration among AI and machine learning engineers, cognitive neuroscience experts, government officials, and psychologists to conduct further research in the future finding solutions for the issues highlighted and ensure the chip's safety and ethical use for everyone.

References

1. Burwell, S., Sample, M., & Racine, E. (2017). Ethical aspects of brain computer interfaces: a scoping review. *BMC Medical Ethics*, *18*(1). https://doi.org/10.1186/s12910-017-0220-y
2. Embattled Giant Of Psychology Speaks His Mind. (1987). *The NewYork Times*. https://www.nytimes.com/1987/08/25/science/embattled-giant-of-psychology-speaks-his-mind.html#:~:text=Skinner%20was%20asked%2C%20would%20he,Skinner%20said
3. Engineering Ethics of Neuralink Brain Computer Interfaces Devices. (z.d.). *Medwin Puplishers*.https://www.researchgate.net/profile/Akram-Jawad/publication/349506653_Engineering_Ethics_of_Neuralink_Brain_Computer_Interfaces_Devices/links/6034302f299bf1cc26e450f1/Engineering-Ethics-of-Neuralink-Brain-Computer-Interfaces-Devices.pdf
4. *Exclusive: Musk's Neuralink faces federal probe, employee backlash over animal tests*. (2022, 6 december). Reuters. https://www.reuters.com/technology/musks-neuralink-faces-federal-probe-employee-backlash-over-animal-tests-2022-12-05/
5. Fadziso, T. (2020a). Why Neuralink will Change Humanity Forever? *Asian Journal Of Humanity Art And Literature*, *7*(1), 69–78. https://doi.org/10.18034/ajhal.v7i1.518

6. Good Morning America. (2024, 17 mei). *Sitting down with Neuralink's 1st brain chip implant patient* [Video]. YouTube. https://www.youtube.com/watch?v=tDTr252Xskg
7. Ienca, M., & Ignatiadis, K. (2020). Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges. *AJOB Neuroscience*, *11*(2), 77–87. https://doi.org/10.1080/21507740.2020.1740352
8. Miśkiewicz, J. (2019). THE MERGER OF NATURAL INTELLIGENCE WITH ARTIFICIAL INTELLIGENCE, WITH a FOCUS ON NEURALINK COMPANY. *Virtual Economies*, *2(3)*.
9. Müller, O., & Rotter, S. (2017). Neurotechnology: Current Developments and Ethical Issues. *Frontiers in Systems Neuroscience*, *11*. https://doi.org/10.3389/fnsys.2017.00093
10. *Neuralink's Animal Experiments Under FDA Fire*. (2024). QMed. https://www.mddionline.com/digital-health/neuralink-animal-experiments-again-under-fda-fire
11. Nijboer, F., Clausen, J., Allison, B. Z., & Haselager, P. (2011). The Asilomar Survey: Stakeholders' Opinions on Ethical Issues Related to Brain-Computer Interfacing. *Neuroethics*, *6*(3), 541–578. https://doi.org/10.1007/s12152-011-9132-6
12. Physicians Committee for Responsible Medicine. (z.d.). *Elon Musk's cruel brain experiments*. Physicians Committee For Responsible Medicine. https://www.pcrm.org/ethical-science/animals-in-medical-research/cruel-brain-experiments

13. Rosenfeld, J. V., & Wong, Y. T. (2017). Neurobionics and the brain–computer interface: current applications and future horizons. *Medical Journal Of Australia*, *206*(8), 363–368. https://doi.org/10.5694/mja16.01011
14. Steinert, S., & Friedrich, O. (2019). Wired Emotions: Ethical Issues of Affective Brain–Computer Interfaces. *Science And Engineering Ethics*, *26*(1), 351–367. https://doi.org/10.1007/s11948-019-00087-2
15. Torrella, K. (2022, 11 december). Neuralink shows what happens when you bring "move fast and break things" to animal research. *Vox*. https://www.vox.com/future-perfect/2022/12/11/23500157/neuralink-animal-testing-elon-musk-usda-probe
16. *US FDA clears Neuralink's brain chip implant in second patient, WSJ reports*. (2024, 20 mei). Reuters. https://www.reuters.com/science/us-fda-clears-neuralinks-brain-chip-implant-second-patient-wsj-reports-2024-05-20/
17. Zhao, Z., Nie, C., Jiang, C., Cao, S., Tian, K., Yu, S., & Gu, J. (z.d.). Modulating Brain Activity with Invasive Brain–Computer Interface: A Narrative review. *Brain Sciences*, *13*(1), 134. https://doi.org/10.3390/brainsci13010134