

Remote Controlled Individuals? The Future of Neuralink: Ethical Perspectives on the Human-Computer Interactions

Maria Cernat¹, Dumitru Borțun¹ and Corina Sorana Matei²

¹*Communication and Public Relations Department, National School of Political Science and Administration, Romania*

²*Communication and Public Relations Department, Titu Maiorescu University, Romania*

Keywords: Human-Computer Interaction, Political Economy of Scientific Discoveries, Ethics of Computer Science Research, Commodification of Scientific Discoveries.

Abstract: In an experiment presented to the public, a monkey with a Neuralink inserted in its brain is able interact directly with the computer. The Neuralink experiment opens the door to an extremely complex debate with questions ranging from ontology to epistemology. We explore the political economy of these cutting-edge technologies. What we aim to investigate are ethical questions, namely: is it ethical to install such a device in someone's brain and connect it to a computer? And who controls the computer, since it is plausible to assume that the communication could be bidirectional? We argue that this is in fact the key question we, as social scientists, IT specialists, and computer science specialists, have to ask and attempt to find answers to. Oftentimes, scientific discoveries could lead to disasters and in the era of "surveillance capitalism" we could easily imagine a scenario where companies are competing to gain access to our consciousness, and where our decisions are being marketed and sold to the higher bidder. Scientific discoveries do not occur in a purely rational society and questions of power, access, and control are vital for a future where technology and society are not at odds.

1 INTRODUCTION

The purpose of this article is to pinpoint certain trends and dangers that may emerge from the development of increasingly advanced technologies that optimize human-machine interaction to the point where our brains can communicate directly with intelligent machines. In October 2021 the journal *Nature Medicine* published an article about how a severely depressed patient was treated for severe depression with a chip implanted and connected directly to her brain. "We developed an approach that first used multi-day intracranial electrophysiology and focal electrical stimulation to identify a personalized symptom-specific biomarker and a treatment location where stimulation improved symptoms. We then implanted a chronic deep brain sensing and stimulation device and implemented a biomarker-driven closed-loop therapy in an individual with depression. Future work is required to determine if the results and approach of this n-of-1 study generalize to a broader population." (Scangos, Khambhati, Daly et alii, 2021)

It was for the first time that scientists have attempted to treat a psychological illness such as depression using mechanistic methods. Interestingly enough, the article fails to tackle ethical issues. The Food and Drug Administration's standards are said to have been applied to NeuroPace RSN (responsive cortical stimulation for the treatment of refractory partial epilepsy). In other words, their research was guided by the standards already available for technology used in epilepsy. (Sun, Morell 2014). The study only set out to show that such treatment is possible and that the treatment of mental illness from a mechanistic perspective would be possible:

"In this study, we established proof-of-concept for a new powerful treatment approach for neuropsychiatric disorders. The new framework presented in this article could advance biomarker-based neural interfaces and enhance the mechanistic understanding and treatment of a broad range of neuropsychiatric conditions." (Scangos, Khambhati, Daly et alii, 2021).

Researchers do not question whether it is desirable or ethical at the same time. Just whether it is possible. This is why we believe our work is of vital

interest in the field of human-machine interaction discussions. The fact that some things are possible does not automatically make them desirable or morally acceptable. Success in one patient's course of treatment can also pave the way for reprehensible uses of technology. **If the competent authorities have approved this technology for epilepsy, perhaps the ethical consequences derived from its use to treat depression are different.**

More ethical issues are "emerging" in the field of emerging technologies, as they are highlighted by SIENNA Project (*Sienna Report*, 2020), an international research program on these technologies, under the auspices of the European Commission, coordinated by the University of Twente, the Netherlands. Some of the moral, as well as social, economic, and political issues concerning the Human Enhanced Technologies (HET) are:

- Right to privacy; the intrusion and the monitoring of man's life by HET, including the physiological intimacy is quite likely in the future;
- Freedom to be "imperfect" refers to the charm and the uniqueness of someone's personality which could be rendered by his/her own imperfection;
- Addiction to emerging technologies affects individuals regardless of their age, but HET and Artificial Intelligence could only deepen this trend, invalidating man as social being, as professional and even as rational being.
- Social inequalities and discrimination would deepen the differences in social status, financial power, inequality of chances between various social categories.
- Misuse of emerging technologies by transforming them into tools or weapons in order to achieve immoral or illegal purposes (theft, violence, illegal surveillance, espionage, etc.).
- Ownership and censorship of expensive HET and AI devices; these issues are linked to the economic and financial problems posed by the sale or loan of technologies to users, which would allow manufacturers to enforce certain restrictive conditions and increase addiction to their products.
- Problems of security, safety and liability refer to the risk that HET and AI would trigger multiple types of errors and accidents, especially where they are supposed to interact with humans or replace humans endowed with reason, intuition, instinct and wisdom.

- Possible weaponization of enhancements; It is known that, throughout history, the military field was the first to benefit from technological innovations and inventions which, after their wear and tear, were upgraded and disseminated to the general public.

Technology is an integral part of our lives. Now we have smartwatches that can even detect sleep quality, smart home lighting systems, remote-controlled alarm systems, phones with dozens of apps. We have reached the point where it is no longer enough to wonder whether technology would become part of our lives, but when it will become part of our being. Maybe it is time to wonder if the future reshaping of human nature would predictably lead to a new form of alienation, by amputating man's ability to decide autonomously, to exercise his/her free will and his/her reasoning.

Because technology by itself is meaningless, we need to ask how it is used. Shoshana Zuboff in her wide-ranging work "Surveillance Capitalism" points out that technology is currently used to generate profit, while the raw material is our lives and our personal data. (Zuboff 2019)

However, technology is still just part of our lives, not our being. Still, the interaction between humans and machines is mediated by thoughts, intentions, and language. What will happen when technology still operates according to the logic of profit becoming, at the same time, part of our being not only of our lives?

In 2017 The Guardian journal published an article in which famous inventors of equally famous technologies said that they themselves no longer wanted to use them! In the extensive article signed by Paul Lewis heavy names in the IT industry such as Justin Rosenstein, the engineer who invented the "like" button, Tristan Harris - a former Google employee with a psychology degree from Stanford and the first one to write a famous memo about Google's unethical products — or Loren Brichter, the designer who created Twitter's notification checking mechanism were horrified by the effects of the technologies they had created.

Now, if we have such reactions when technology is part of our lives, not part of our being, how can we assume that things will go in the right direction? And as long as we put emphasis on profit, how could it be otherwise? Here's what Tristan Harris says: *"tech companies never deliberately set out to make their products addictive. They were responding to the incentives of an advertising economy, experimenting with techniques that might capture people's attention, even stumbling across highly effective design by accident."* (Lewis 2017).

Chris Marcellino, who, along with Justin Santamaria, is the designer of Apple's patent for "managing notification connections and displaying icon badges", makes it clear that the fault lies with the system: "It is not inherently evil to bring people back to your product. It's capitalism." (Lewis 2017)

The point is not that scientists are amoral geniuses. The point is that we promote amoral geniuses as a model for success in a system where profit is all that matters. And that we don't even offer the chance for an ethical discussion before creating technologies that are going to change people's lives for decades. Here's what Loren Brichter, the designer who created Twitter's notification checking mechanism, says: *"Pull-to-refresh is addictive. Twitter is addictive. These are not good things. When I was working on them, it wasn't something I was mature enough to think about. I'm not saying I'm mature now, but I'm a little bit more mature, and I regret the downsides."* (Lewis 2017). We can learn enormously from these statements. First and foremost, ethical concerns ought to be in the limelight of all technology-related conferences, in college departments training future engineers, in research grant applications, in the evaluation of grants.

2 POSTHUMANISM AS A RE-EXAMINATION OF THE HUMAN-OBJECT RELATIONSHIP

The phrase "posthumanism" has emerged in connection with the development of artificial intelligence and, more recently, the robotization of thought and communication. It expresses philosophers' concern about the dethronement of Man from the position of "supreme value of the known universe", the center of all endeavors to know, evaluate and transform reality - the ancient status of Man, the "measure of all things" according to Protagoras (Plato 1989, 181-299). Is this concern reasonable? Are there enough data to suggest the "end of humanism", or, on the contrary, does the new status of the human being rather oblige us to rethink the values of humanism, to shape a more realistic, active, and effective humanism, as a complement to the humanism of the past, which has been rather narcissistic, rather contemplative and rather powerless?

In other words, we should ask ourselves whether a humanism without anthropocentrism is possible, a humanism that preserves "man as the supreme value"

but abandons "man is the measure of all things". Anthropocentrism has had two major consequences: 1) it has devalued God, who is no longer at the center of the world, the place occupied by Man, and who, instead of creator becomes creature: it was not god who created man, it was man who created god (Feuerbach) ; 2) which led to the dehumanization of animals and objects - the main worldly instances against which humanists defined the deminuteness of being human: from "I am happy to be human and not animal", as Empedocles boasted (Laertios 1997, VI, 80), to "humanity (. ...) never only as a means, but always at the same time as an end in itself" (Kant 2007, 324). Just as ecosophy proposed the rethinking of the relationship between man and animal, posthumanism proposes the rethinking of the relationship between man and object - specifically, the relationship between man and robot; generically, with artificial intelligence, as objectified humanity.

Humanism does not exist as such, but as a plurality of humanisms; in the twentieth century there was talk of speculative and scientific humanism, rationalist and religious, existentialist or Christian, evolutionary or aesthetic - finally, of real, revolutionary or integral humanism. In the 21st century, there is talk of a new perspective on the human condition: posthumanism. The "dialogue of humanisms" (Garaudy 1960) has hitherto been conducted as a confrontation of humanisms. The collapse of the "Berlin Wall" has enabled us to see that the lines of the struggle for a better society and for the safeguarding of civilization are not those drawn during the Cold War. And the COVID 19 pandemic has shown us that the true lines of approach separate the manifestations that lead to the full development of human capacities, which promote Life, and those that paralyze the exercise by billions of people of their essential human faculties, which stifle Life.

To save the future, a paradigm shift is needed, enabling the directions of development of contemporary society to be understood in new, contemporary terms, and not in the terms in which the former problems of the human condition were formulated. In this context, we believe that the debates on "posthumanism" are flawed by two shortcomings: on the one hand, a lack of topicality and, on the other, the absence of the ethical referential proposed by the Judeo-Christian paradigm, which preserves the pride-of-being-man but rejects anthropocentrism.

The authors of the paper try to show that robotization and hyper-technologization do not endanger humanism per se, but a certain historical

form of it, namely the ideological or "rhetorical humanism" (Uscătescu 1987, 40). This paper shows that framing the discussion on "transhumanism" within the old paradigm, which was built around "man as a rhetorical element" (Uscătescu 1987, 41) is tantamount to framing it in the midst of ideological alienation; alienation entails "a new metaphysics" (Max Scheler), without which we will not be able to understand the current technological progress and, consequently, will not be able to take advantage of its good fruits, just as we will not be able to avoid biting into its poisonous fruits.

We see a return to dialogue as a *sine qua non* for forging a broad alliance in the future against anti-humanist tendencies, against forces which, by their actions, demean man and endanger his life and the civilization he has created. Three decades after the fall of the Wall between East and West, it has become clear that the decisive battle for the future of civilization is being fought on the barricade that separates contempt for man and his life from love for what he represents and could represent - in short, contempt for life from "respect for life" (Albert Schweitzer).

This question thus arises: is this an instance of the struggle between humans and robots - in other words, between the human in us and the technology created by some of us? Does the creature come to dominate the creator to the point where the creator has to fight his creature? So far, we have been told that Pygmalion falls in love with his creation, being left to admit that he falls in love with himself, with a Self objectified by creation (Pygmalion as a disguised Narcissus!). In George Uscătescu's view, saving European culture, regaining its unity, means harnessing its deepest idea, which predates humanism and the human sciences: the idea of Freedom. The problem of culture is the problem of freedom or, more precisely, it is the problem called by Paul Hazard "the eternal conflict between authority and freedom" (Hazard 1973, p. 79), if it is not that which can be formulated, even more precisely, in terms of Andre Glucksmann: the opposition between the real freedom of the "plebs" and the "authority of knowledge" which can give rise to an ideology definable as "the science of authority, the science of the set of methods, ideas and behaviors which allow the conquest, preservation, and consolidation of power in the twentieth century" (Glucksmann 1991, 59). This perspective, open to Culture and Freedom, implies a new philosophy, a new "metaphysical consciousness" (Uscătescu 1987).

3 MAKING THE POSTHUMAN POSSIBLE

Neuralink is a famous company in which the famous Elon Musk has invested heavily to produce revolutionary technologies that lead to better communication between brain and computer. At first glance, everything seems harmless and very promising. People suffering from paralysis will be able to interact seamlessly with a computer, which will improve their lives. Why should such praiseworthy initiatives be seen as anything other than technological advances?

For many reasons, of which the first and most argued concern the so-called post-humanism. Posthumanism is a philosophical concept whose origins stem from the philosophy of Nietzsche, according to some scholars. Post means something that transcends, that follows. In this case it also indicates something that challenges humanism, essentially the idea that humans are exceptional and that there is a fundamental difference between man and nature, man and animal, man and... machine.

As early as 1999, Katherine Hayles, a professor of English at the University of California, published a seminal work in the field, "How We Became Posthuman: Virtual Bodies in Cybernetics, Literature and Computer Science", in which she attempted to explore the main consequences of our increasingly close interactions with technology and, more importantly, the way in which technology becomes part of our very being (Hayles 1999).

Recent research in the field of human-machine relations is likely to bring to light a very complex and explored issue in epistemology (how much can we know with the help of machines, how can knowledge be enhanced with the help of technology), ontology (who am I once my consciousness is "downloaded" into a computer) and, of course, ethics (who controls my virtual self, who controls the network into which I download my experiences, who controls the eventual reverse flow of information from the machine to my brain).

In her paper, Katherine Hayles explores some of these directions. She starts from the problematic mind-body dualism to show that the obsession with control, with the separation of a spiritual, immaterial 'essence' from the immanent, decaying body, has populated the philosophical imagination for many centuries. The theme originally developed in a systematic way by René Descartes acquired major importance in the 20th century in the so-called "philosophy of mind" branch of philosophy. The nature of consciousness, how it exists and the

question of whether it can be reduced to materiality, to the states of the brain, gave rise to extremely heated debates. On one side of the fence we have those who argue that consciousness cannot in fact exist under any circumstances independently of the body, of our brain. Obviously, on the other side, we find those who argue that all our thoughts, all our inner feelings, are an expression of brain activity and that they do not exist beyond them.

It is interesting how the idea of 'downloading' our inner experiences into a common consciousness, what Slavoj Žižek calls Singularity (i.e. a single common consciousness), shows a potential contradiction between these positions in the philosophy of mind. On the one hand, empirical research funded by Elon Musk goes implies that human consciousness is indeed nothing more than the result of the functioning of the brain. Neuralink's goal is to identify the specific brain activity that is responsible, for example, for the desire to move an object on the screen, and then to translate that nervous input from the brain into commands to the computer.

The question is whether in the future it will not only be the intention to move an arrow on the screen that will be read by the computer. Perhaps our entire consciousness will be "downloaded" into a computer. This starts from the second premise - that the states of our consciousness are completely reducible to the states of the brain - and demonstrates the possibility of the first - that our entire consciousness could exist outside these states. In other words, the imagined experiment of transferring not just simple commands, but all consciousness into a computer, shows that it is reducible to brain states, but also translatable into electrical signals that the computer can interpret.

It is clear that if we accept that we could "download" the consciousness into a computer we are already relying on a hidden assumption. Namely, that consciousness actually exists outside the body, in the mind, not in the body. The whole argumentative framework rests on this idea. That we can separate the mind from the body. But not from any part of the body, but from everything that is not the brain. In other words, we can separate the mind from a part of the body and locate the mind in the brain.

Division and control are two other premises that can justify these attempts to locate the mind and separate it. In European culture in particular, the relationship with the body is a very problematic one. The body is seen as fallen, corruptible, weak, shameful. Since forever, monks have been engaging in extreme self-punishing rituals in order to control their bodies. We can also see here the desire to dominate nature and the hatred of women. They are

both considered an expression of nature and, as a result, a source of corruptibility. This attitude begins with hatred of one's own body seen as sinful, flawed, imperfect and, above all, uncontrollable.

The idea of "unloading" consciousness into a machine is the emanation of this type of culture in which the body is either a source of corruptibility or an imperfect "package" of a higher-order essence that deserves to be transposed and preserved in an inorganic shell, such as that of the machine, and, above all, controllable!

Another assumption on which the simplistic illusion of the discharge of our consciousness is based is pointed out by Katherine Hayles. "As I have repeatedly argued, the human being is primarily the embodied being, and the complexities of this embodiment translate into the fact that, in fact, human consciousness manifests itself in very different ways from intelligence downloaded into cybernetic devices." (Hayles 1999, 284) Thus, the existence of consciousness as a set of mental states that can be translated into computer states is, beyond the obvious empirical limitations, impossible to put into practice due to other types of reasons. Consciousness is the result of being in a body. The very idea of separating and controlling both mind and body can be understood as the obsession of a white man whose wealth allows him to follow his urges, sometimes taking them to unsuspected heights (of the ridiculousness). Harder to understand might be the lack of reactions questioning the very fact of having so much financial power in the hands of a single individual who can act despotically and arbitrarily without being held accountable for his actions. But, until that point is reached, we have yet to examine the theoretical field of posthumanism as a theory that comes to encompass the results of such a potential dissolution of the human-machine boundary.

In 1950 the famous mathematician Alan Turing published an article in *Mind* magazine that has shaped the research and intrigued the minds of philosophers and computer scientists for decades. In this article, Alan Turing, in the section of the article entitled *The Imitation Game*, proposes a mental experiment. The mathematician says that the question "can machines think?" was being asked more and more insistently. As it was not possible to answer by a simple opinion poll, he proposed the following experiment.

A, a man, B, a woman, and C, the questioner, took part. C has to talk in writing to A and B without seeing them and has to figure out which is the man and which is the woman. Then we have to imagine what would happen if A's place was taken by a car? Could C, the questioner, the one who talks in writing to the

subjects he cannot see, figure out which is the man and which is the car? "Now we ask the question: <What will happen when a car takes A's role in this game> Will the questioner make as many mistakes in correctly estimating roles as if the game were played between a man and a woman? These questions replace the question, <Can machines think?>" (Turing 1950, 433).

4 CONCLUSIONS

As interesting as these debates are, they suffer from two major drawbacks. First, there are serious empirical limitations that turn the supposed downloads of life experiences into a computer into pure Elon Musk marketing strategy. The idea that you may have had a wonderful sexual experience that you want to download to your computer to relive at a later time or, if exhibitionism doesn't get in the way and there is no moral burden of intimacy with your partner, to share with others, is, for now, at best an emanation of a mind too deeply and too early immersed in the pornographic ocean. Secondly, these discussions seem to totally bypass power relations, money, and its distribution.

The political economy of the posthuman refers to questions in the area of ethics which it transfers to the political plane. If the technology that will enable cranial implants designed to cure paralysis becomes available, who will have access to it? If Elon Musk decides that perhaps women should not have access to it, who could compel him to act morally? After all, is it his money? He has the right to select his clientele. Let's say in the case of women it would be too blatant. But what if his despotism were to be diverted to more problematic categories? If, for example, he decided that drug addicts, socially downgraded should not have access to it, who would stop him?

In his latest book, *Hegel in a Wired Brain* Slavoj Žižek discusses Musk's proposed experiments at length. He's not catastrophic, apocalyptic or promoting moral panics. But he asks the right questions. Now we see a monkey thinking about moving the cursor on the computer screen and the computer reads his intentions directly from his brain. But what would it be like to have reverse communication? Has the computer sent commands to the monkey?

This is where things get really interesting. Also in the 1950s, at the height of Cold War hysteria, psychiatrist Donald Cameron conducted a series of terrifying experiments on his patients. Obsessed with the idea that the Russians could program an ordinary

human to act on command at a moment's notice, the Americans invested heavily in psychological programming techniques. The harshest treatments, electric shocks, huge doses of psychotropic drugs, torture techniques, sleep deprivation, sexual abuse, endless repetition of sound sequences - nothing stood in the way of Dr. Donald Cameron's mission to re-program schizophrenia patients. It was only a happy accident that in 1973, when Cameron's sinister CIA-funded MKUltra program in Canada came to an end, the documents that documented its existence were not destroyed. They had been misplaced in the financial section. So, in 1977, the US Congress was able to bring members of Donald Cameron's team to testify.

In his latest book, Slavoj Žižek addresses precisely this problem of the possibility of programming, of transmitting orders remotely to a human brain connected to Neuralink. In his many lectures, Žižek asks the natural question: is it OK that the monkey communicates so well with the computer? But who owns the computer? And, if Žižek is to be believed, the great danger will come not just from the fact that we might receive commands from a computer once we have the cranial device implanted, but that we will not realize we are receiving them. Because, in his view, the totalitarianism of an authoritarian government like China's is still not as dangerous as a potential system where you don't even know you're being controlled.

REFERENCES

- Hayles, Katherine. 1999. *How We Became Posthuman. Virtual Bodies in Cybernetics, Literature and Informatics*. Chicago, London: University of Chicago Press.
- Feuerbach, Ludwig. 1961. *Esența creștinismului*, București: Editura Științifică.
- Kristof, Nicolas. 2020. The Children of Pornhub. Why Does Canada Allow this Company to Profit of Videos of Exploitation and Sexual Assault? *The New York Times*.
- Garaudy, Roger. 1969. *Marxismul secolului XX*, București: Editura Politică.
- Glucksmann, André. 1991. *Bucătăreasa și Mâncătorul de oameni*. București: Humanitas.
- Hazard, Paul. 1973. *Criza conștiinței europene. 1680-1715*. București: Univers.
- Kant, Immanuel. 2007. *Întemeierea metafizicii moravurilor*, București: Humanitas.
- Laertios, Diogenes. 1997. *Despre viețile și doctrinele filosofilor*, București: Minerva.
- Lewis, Paul. 2017. Google, Twitter and Facebook workers who helped make technology so addictive are disconnecting themselves from the internet. Paul

- Lewis reports on the Silicon Valley refuseniks alarmed by a race for human attention. The Guardian, <https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia>
- Platon. 1989. Theaitetos, în vol. *Opere VI*, București: Editura Științifică și Enciclopedică.
- Project MKUltra, The CIA's Program of Research in Behavioral Modification*, U.S. Government Printing Office, Washington, 1977.
- Scangos, K.W., Khambhati, A.N., Daly, P.M. et al. Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nat Med* 27, 1696-1700 (2021). <https://doi.org/10.1038/s41591-021-01480-w>
- SIENNA Report, 12.05.2020, section "D3.4: Ethical Analysis of Human Enhancement Technologies — WP3 Human Enhancement", pp. 77-88, file:///C:/Users/user/Downloads/Attachment_0%20(1).pdf
- Sun, F. T. & Morrell, M. J. The RNS System: responsive cortical stimulation for the treatment of refractory partial epilepsy. *Expert Rev. Med. Devices* 11, 563-572 (2014).
- Turing, Alan. 1950. Computing Machinery and Intelligence, *Mind*, vol.LIX, nr.236.
- Uscătescu, George. 1987. *Proces umanismului*. București: Editura Politică.
- Zuboff, Shoshana. 2019. *Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, New York: Public Affairs.
- https://www.nytimes.com/packages/pdf/national/13inmate_ProjectMKULTRA.pdf
- Žižek, Slavoj, *Hegel in a Wired Brain*, Londra, New York: Bloomsbury Academic, 2020.