Control of AI or control over AI

Conference Paper · January 2023

DOI: 10.1063/5.0165266

V. V. Zinchenko

National Academy of Pedagogical Sciences; Ukrainian Institute of Arts and Scienc…

68 PUBLICATIONS 186 CITATIONS

SEE PROFILE

Mykhailo Boichenko

National Taras Shevchenko University of Kyiv

43 PUBLICATIONS 69 CITATIONS

SEE PROFILE

RESEARCH ARTICLE | NOVEMBER 09 2023

Control of AI or control over AI

V. Zinchenko; M. Boichenko; V. Levkulych; Z. Samchuk; O. Polishchuk; M Debych

Check for updates

AIP Conf. Proc. 2948, 020002 (2023)
https://doi.org/10.1063/5.0165266

Control of AI or Control over AI

V Zinchenko[1, a)], M Boichenko[2,b)], V Levkulych[3,c)], Z Samchuk[4,d)], O Polishchuk[5, e)] and M Debych[1, f)]

[1]Institute of Higher Education of the National Academy of Educational Sciences of Ukraine, Kyiv, Ukraine
[2]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[3]Uzhhorod National University, Uzhhorod, Ukraine
[4]Kuras Institute of Political and Ethnic Studies of the National Academy of Sciences of Ukraine, Kyiv, Ukraine
[5]Khmelnytskyi Humanitarian-Pedagogical Academy, Khmelnytskyi, Ukraine

[a)] Corresponding author: vvzinchenko@ukr.net
[b)] boichenko.m.@knu.ua
[c)] professor@feo.ua
[d)] zsamch@ukr.net
[e)] Olenkapol@gmail.com
[f)] professordfn@gmail.com

Abstract. The technological development of mankind is significantly ahead of its institutional development. This necessitates the revision of existing means of institutional control over AI. Revolutions in the development of computer science (the flourishing of machine learning, artificial intelligence) and in the development of biology, in particular neuroscience open up new opportunities in the AI implementation and embodiment. Scientists invent biometric sensor as a direct link between AI and the human brain. This and other scientific discoveries make it necessary to consider the following questions: the risk of the emergence of digital dictatorships; the potential for AI to control or serve surveillance capitalism; AI as a possible defense against the misuse of AI. The methodology of this research is value analysis and institutional analysis. The aim of the article is to outline the current situation with the definition of possible value criteria for determining the goals of AI control and possible institutional ways of implementing such control. It is provided an assessment of the current effectiveness of existing institutional controls over AI.

INTRODUCTION

Artificial intelligence (AI) opens up wide opportunities for more effective use of human intelligence and gives a new impetus to the project of Enlightenment – significant progress of society thanks to intelligence. Together with AI, humanity gains previously unknown prospects for social changes and changes not only in the external environment, but also in human's internal organic structure. Human evolution begins to critically depend on the efficiency and reasonableness of human use of AI [15]. At the same time, a human being, together with AI, acquires a new quality: his opportunities to control himself, his activities, and, accordingly, the results of these activities, increase. But part of this control will belong to AI. Who will determine the share of control of the AI – the AI itself or the humans? The answer to this question depends partly on the technical characteristics of AI, but partly determined by human values and goals. Let's try to outline these boundaries.

RESEARCH METHODOLOGY

The methodology of this research is value analysis and institutional analysis. Both scientific articles and analytical publications in the media will be considered – with the aim of revealing not only the conceptual foundations of the

IV International Scientific Forum On Computer And Energy Sciences (WFCES II 2022)
AIP Conf. Proc. 2948, 020002-1–020002-6; https://doi.org/10.1063/5.0165266
Published by AIP Publishing. 978-0-7354-4732-5/$30.00

study of the problem of control over AI, but also the main directions of public opinion formation regarding the need for such control. An assessment of the current effectiveness of existing institutional controls over AI will also be provided.

The aim of the article is to outline the current situation with the definition of possible value criteria for determining the goals of AI control and possible institutional ways of implementing such control.

BIOMETRIC SENSOR AS A DIRECT LINK BETWEEN AI AND THE HUMAN BRAIN

We are currently facing two simultaneous revolutions: the development of computer science (the flourishing of machine learning, artificial intelligence) and the development of biology, in particular neuroscience. This gives us the necessary understanding of how the human brain works. One and a half centuries of biological research can be summed up in three words: organisms are algorithms. And we are now learning how to decipher these algorithms.

For example, in 2021, researchers have achieved that a brain implant relieves depression. Autonomous electrical stimulator suppresses bouts of depression in the emotional centers of the brain.

In severe cases, depression is not helped by drugs or electroconvulsive therapy, no matter in what combinations they are used. Such patients can only hope for the latest treatments, such as those developed by researchers at the Weill Institute of Neurosciences at the University of California, San Francisco. They created an implant for electrical stimulation of the deep layers of the brain: extremely thin electrodes are inserted into a certain area of the brain and stimulate this area with weak discharges [8].

Neuroscientists have long been studying the possibilities of such electrical brain stimulation for the treatment of various neuropsychiatric diseases, from depression to drug addiction. But in the case of depression, such studies usually end in different ways – the effect is either there or not. The authors of the work believe that the thing is that usually the implant works all the time. In their experiments, everything was different: the implant read the activity of the brain and, using a special algorithm, guessed the approach of an attack of depression – accordingly, the implant sent antidepressant impulses when it was needed. However, from a neurobiological point of view, depression is quite diverse. Our mood and emotions depend on several neural centers, and depression in different people arises due to malfunctions in different neural networks. That is, an individual approach is needed here, and it is precisely this individual approach that the researchers showed with one patient with severe clinical depression. The tests lasted ten days, and it turned out that the whole thing was in the signals that went to the amygdala from the area on the border of the so-called lower capsule and the lower zone of the striatum. To place the implant, a hole 1.5 mm in diameter was made in the skull, through which electrodes were inserted into the amygdala and into the aforementioned region between the lower capsule and the lower zone of the striatum. An electrode in the amygdala monitored when it would begin to generate depressive signals, and at that moment another electrode stimulated the aforementioned border area to bring the amygdala back to normal. During the day, the patient received about 300 stimulations, which lasted a total of about half an hour. In fact, the implant works as a cardiac pacemaker, which helps the heart muscle not to stray from the correct rhythm – only here, instead of the heart, there will be one or another area of the brain prone to depressive activity. The patient really got rid of depression, at least from especially severe attacks. She was no longer tormented by suicidal thoughts, her mood improved, and in general her life went smoothly from the moment she was given an electrically stimulating implant - and this happened last summer. The implant is powered by a battery that is built into the skull bones and should last for 10 years (so there is no need for too frequent operations to replace a dead battery) [8].

Stimulation is not everything, and the patient will need to work with psychotherapists, but with an implant, she will at least be able to follow their recommendations. So far, this is the first and only example of such a treatment for depression, but there is every reason to believe that a similar approach will work for other patients [6].

Elon Musk announced that in 2022 his company Neuralink plans to implant chips in people [3]. The chip, which Musk's company will implant in people, is designed to record and stimulate brain activity. It will be used in medicine to treat serious spinal cord injuries, neurological disorders and to help the paralyzed. The Neuralink chip has performed well in monkeys and can be safely removed from the body, Musk said. Next year, Neuralink is awaiting approval from the US government regulator, the Food and Drug Administration.

Perhaps the most important invention for both revolutions is the biometric sensor, which translates the biochemical processes in our bodies and brains into electronic signals that a computer can analyze. By deciphering these algorithms, you can create creatures that will be better than people. But how these creatures will subsequently apply these technologies, to be honest, few people have any idea.

THE RISK OF THE EMERGENCE OF DIGITAL DICTATORSHIPS

Another risk of uneven adoption of technology is the emergence of digital dictatorships that can constantly monitor all their citizens. This risk can be condensed into perhaps the most important equation of the 21st century: B*C*D=AHH, where B is biological knowledge, C is computing power, D is data, and AHH (ability to hack humans) – the ability to "hack" each individual person. If you have all the arguments for this formula, "hack" the body, brain, consciousness of a person and understand him better than he himself. «Biological knowledge multiplied by computing power multiplied by data equals the ability to hack humans, ah» [1].

A formula-based system would be able to predict, manipulate, and manipulate a person's feelings and decisions, and ultimately decide for them. In the past, many tyrants dreamed of such a system, but none of them had enough ideas about the biological nature of man, advanced technologies and an array of data available today. We are no longer some kind of mysterious souls, we are animals that can be hacked.

At the World Economic Forum in Davos in 2020, Israeli scientist Yuval Noah Harari spoke about who will control life on Earth in the future, and how people will change in the context of constant technological revolutions.

“Yes, all this information can be used for good, for example, to create an effective healthcare system. But imagine that our arguments fall into the hands of a conditional new Stalin – then we are all threatened with a dictatorship never seen before. But there are already candidates for the role of this new Stalin – let's think about what North Korea will be like in twenty years? Citizens there will be able to put on biometric bracelets that measure all the data, and when a person listens to the speech of the supreme leader on the radio, the “big brother” will know what he really thinks about the ruler. You can smile and clap your hands, but if you are angry at this moment, then tomorrow morning you will find yourself in the Gulag [4].

Harari is a medievalist historian and bestselling author of Sapiens: A Brief History of Humankind, Homo Deus: A Brief History of Tomorrow, and 21 Lessons for the 21st Century [9]. In «Sapiens», he explored our past. In «Homo Deus», he looked to our future. Now (in «21 Lessons for the 21st Century»), one of the most innovative thinkers on the planet turns to the present to make sense of today's most pressing issues. «How do computers and robots change the meaning of being human? How do we deal with the epidemic of fake news? Are nations and religions still relevant? What should we teach our children?» [4].

The algorithm tracks your eye movements, pressure, brain activity and tells you who you are. Even if you hide your orientation from friends and colleagues, Amazon, Google, Facebook can get this data. Algorithms tracking you will tell Coca-Cola, for example, that if the company wants to sell you a drink, you should not show ads with naked girls. You won't even notice it, but corporations know full well that this information is worth billions [16].

Take, for example, North Korea: people there will wear, say, special biometric bracelets. When a person enters the room and sees a portrait of another great leader, the bracelet counts his emotions, pressure and transmits the data to the relevant authorities - this is how a digital dictatorship will look like.

Information control will allow the world's elites to do something even more radical than digital dictatorship. By "hacking" organisms, the elites will be able to reshape the future of life. And this will be the greatest revolution in the history of not just mankind, but all life on Earth. For 4 billion years, the rules for the existence of life on the planet did not change, all living things obeyed the laws of natural selection and organic biochemistry. But now science is replacing evolution by natural selection with evolution by intelligent design. This intention is not of God, but of human being [14].

If this issue is not settled, a tiny group of people, the elite, will gain access to it and will determine the future of life on Earth. Many politicians are like musicians: they play on human emotions and they could have an interest to play on human biochemical system. The politician makes a speech, and the whole country is seized with fear. The politician publishes a tweet, and – an explosion of anger of millions of citizens.

SURVEILLANCE CAPITALISM AND AI

Many researchers note that "surveillance capitalism" is already being introduced ("surveillance/surveillance capitalism" – a term coined by Harvard Business School Professor Emerita Shoshana Zuboff). This also applies to social networks [17]. The main idea of the book comes from the title – we live in an economy in which the surveillance of its participants in the form of constant data collection is an integral part.

“Surveillance capitalism was invented at Google in 2001,” says Shoshana Zuboff in an interview with the magazine Surveillance & Society [18]. According to her point of view, it was in the 2000s that Google realized the importance of user data, which can be used not only to improve user experience, but also to sell advertising. The uninterrupted flow of goods and services in our social networks for Zuboff is not a harmless marketing ploy. On the contrary, it threatens human freedom and autonomy.

Through the collection, possession, and analytics of user behavior data, Silicon Valley companies influence consumer behavior almost imperceptibly, influencing both the choice of a new toothbrush and an election candidate. All existing global regulation, including such progressive projects in terms of protecting user data as the GDPR (General Data Protection Regulation) in the European Union, are not able to stop IT companies from continuing their consistent attack on human autonomy as an opportunity to independently manage their lives.

Zuboff's book has become a concentration of all the fears and dystopias regarding modern technology. In this vision, our personal and collective life ceases to be the space of our freedom and begins to be only a source of data in the course of processes over which we, as citizens and users, have no control. And while the message of the book answers many of the anxieties of civil society, it is necessary to pay more attention to how well surveillance capitalism as an idea captures the complexity of our technological situation.

After the release of the book in American society, the phrase "surveillance capitalism" was picked up by many liberal publicists and public figures, it sounded brightly in the press, and the work itself received mostly only laudatory reviews. Indeed, there seem to be very few people in 2020 who believe that the giants of Silicon Valley are doing everything right. Dissatisfaction with the actions of such companies, which does not happen so often, is shared by both the government and grassroots movements – the US Department of Justice antitrust lawsuit against Google and the public campaign against facial recognition are prime examples of this.

However, the idea of “surveillance capitalism,” which Zuboff traces back to the 2000s, does not take into account the longer history of technology companies and the state in the United States, which began at least in the 1970s. Moreover, many firms already at that time collected data on consumers for sale to the state. For example, TRW even cooperated with the CIA in a similar way. Zuboff prefers to focus on the group of companies in his narrative, ignoring the long historical context in which state and company oversight was carried out.

Many researchers of these processes – and here Zuboff is no exception - in texts about digital technologies are extremely fond of exploiting the "West/China" opposition, where the latter is portrayed as an absolute evil due to the state's strong recourse to technological control in its various variations. But it is enough to deviate somewhat from the idea of a “free market” to see that the state has always existed in the West as well – it simply “hidden” better.

Zuboff does not see capitalism as a complex socio-economic system, she only talks about the history and collection of technology firms. In other words, the question here is how much of a problem it is in specific giants like Amazon, Google, Facebook, or in the very structure of the economy.

The problem of gaining "digital sovereignty" as an opportunity to determine the relationship between technology and political autonomy is acute not only for developing and totalitarian states, but also for Western countries.

AI AS A POSSIBLE DEFENSE AGAINST THE MISUSE OF AI

Officially, the purpose of various forms of control and analysis by AI is to protect the population from potential external threats.

It turns out that some algorithms will take care of our security. They, according to the parameters laid down in them by the developers, will judge who is a danger and who is not. How fair, biased and accurate are these algorithms? Where, by whom and how is this data processed? Who has access to them?

Experts around the world are trying to find answers to these questions. Legislation in different countries reacts very differently to the amazing capabilities of AI surveillance and control systems.

In April 2021, the "Draft EU Regulation on Artificial Intelligence" was presented in the European Union (full name «Commission proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts(COM(2021) 206)» [2].

Experts believe that this document can make the same furore in the legal regulation of artificial intelligence, as did the "European Regulation on the Protection of Personal Data" (GDPR) in 2018. The new proposal to the regulation, in addition to AI, also regulates issues related to the face recognition system.

The "Draft EU Regulation on Artificial Intelligence" establishes different degrees of risk for the application of artificial intelligence. High-risk cases may be prohibited or strictly regulated. For example, high-risk systems include mass surveillance systems that have a face recognition function in public places. An exception is cases of searching for missing persons or situations confirmed by the court related to the prevention of terrorist threats. In other cases, face recognition systems want to be completely banned in the EU.

Also, according to the European Regulation for the Protection of Personal Data, biometric data are personal data obtained as a result of special technical processing that relate to the physical, physiological or behavioral traits of a person. Biometric data can accurately identify a person. This means that the information received from face recognition systems is also biometric personal data, which has its own specifics for processing.

In the United States, there is no federal law yet that would regulate facial recognition issues. However, there have accumulated a lot of court cases, the subject of which is the legitimacy of facial recognition.

For example, there was a lawsuit against Facebook alleging that the company "illegally collected and stored the biometric data of millions of users without their consent." Users from the state of Illinois claimed that Facebook violated their rights with the "tag suggestion" feature on photos. The 9th Circuit Court of Appeals agreed. And considered that "the development of a face template using facial recognition technology without consent ... invades privacy and violates the interests of a person" [7].

A lawsuit was filed in New Jersey against the police for using facial recognition technology to identify a suspect, as a result of which an innocent person was identified and detained [5].

The unjustly accused man spent 10 days in prison. The police did not even use fingerprints for identification, relying only on algorithms. As a photo of the perpetrator, they used a photo on a fake driver's license found in a rented car driven by a real perpetrator.

The US Congress has been trying to pass federal legislation that would regulate issues regarding facial recognition. But several bills in the 116th Congress were never passed into federal law. Including:

- Commercial Facial Recognition Privacy Act of 2019 [10];
- The Ethical Use of Face Recognition Act of 2020 [11];
- The Face Recognition and Biometric Technologies Act of 2020 [12].

Washington, DC, for example, has a law regulating the use of facial recognition services. Its first section states that the unrestricted use of such services by government agencies has social implications that need to be addressed and addressed. How to consider and eliminate such consequences? Legislation is needed to establish appropriate safeguards.

The state should use facial recognition technologies in a way that benefits society. And prohibit uses that threaten democratic and civil liberties. For example, government and local government agencies may use facial recognition technologies to find or identify missing people or to identify the dead.

In the United States, a bill is being prepared "Fourth Amendment Is Not For Sale Act" (the act "The Fourth Amendment Is Not For Sale" [13]). It will clarify the fourth amendment to the US Constitution. The bill would ban law enforcement from buying data from companies like Clearview AI, a startup that has amassed a huge bank of publicly-sourced photos of people.

Government organizations bought data from Clearview AI that would normally require a warrant. Thus, the fourth amendment to the US constitution, which prohibits illegal searches and detentions, was violated. If the law is passed, government agencies will need to obtain a warrant before requesting location data from data brokers.

CONCLUSION

Who should own the information? The discussion about this has just begun. We cannot expect an immediate answer to this important question. After all, the future of not only mankind, but also life itself on the planet depends on the answer to it. But progress has been limited so far. Most people are completely unaware of what is going on and what is at stake. Many governments also shrug it off: we have more urgent business to attend to. And it's very dangerous.

How to manage information ownership? Unlike land and industrial equipment, information is everywhere and at the same time nowhere, it can be copied, it spreads at an incredible speed, and so on. Who owns information about me? Corporations own most of the data these days, and people are worried about that. But empowering corporations and governments to nationalize information leads to a digital dictatorship.

When we have algorithms that can understand us better than we can, they will be able to predict our desires, manipulate our emotions, and even make decisions for me. If we are not careful, the era of digital dictatorship will come.

In the 20th century, democracy replaced dictatorship because it was better at processing data and making decisions. Democracy distributes information and empowers institutions and people to make decisions. Dictatorship, on the other hand, concentrates all information and decision-making in one hand.

However, in the 21st century, the biotechnological revolution may swing the pendulum in the opposite direction: the centralized distribution of information may become more efficient. If democracy does not adapt to new conditions, new people will live under a digital dictatorship.

REFERENCES

1. Betazone Davos 2020. How to survive the 21st century with Yuval Noah Harari. World Economic Forum (2020), https://youtu.be/Rw9FSYH6kL8
2. European Parliament, *Artificial intelligence act* (2021), https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/694212/EPRS_BRI(2021)694212_EN.pdf.
3. Hamilton, Elon Musk said Neuralink hopes to start implanting its brain chips in humans in 2022, later than he anticipated (Dec 7, 2021), https://www.businessinsider.com/
4. Y. N. Harari, *21 Lessons for the 21st Century* (Kindle Edition, 2018).
5. K. Hill, Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match (Dec 29, 2020), https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html
6. R. Marks, Treating Severe Depression with On-Demand Brain Stimulation. UCSF Team Provides Immediate, Long-Term Relief for Patient’s Symptoms (October 4, 2021), https://www.ucsf.edu/
7. Patel v. Facebook, Inc., No. 18-15982 (9th Cir. 2019), https://law.justia.com/cases/federal/appellate-courts/ca9/18-15982/18-15982-2019-08-08.html
8. K. W. Scangos, A. N. Khambhati, P. M. Daly et al., Nature Medicine **27**, 1696-1700 (2021).
9. S. A. Shay, Yuval Harari’s Insights Are As Old, And As Ominous, As The Tower Of Babel (October 28, 2019), https://jewishweek.timesofisrael.com/
10. US Congress, Commercial Facial Recognition Privacy Act of 2019 (2019), https://www.congress.gov/bill/116th-congress/senate-bill/847/text
11. US Congress, The Ethical Use of Face Recognition Act of 2020 (2020), https://www.congress.gov/bill/116th-congress/senate-bill/3284/text
12. US Congress, The Face Recognition and Biometric Technologies Act of 2020 (2020), https://www.congress.gov/bill/117th-congress/house-bill/3907/text?r=7&s=1
13. US Congress, Fourth Amendment Is Not For Sale Act (2021), https://www.congress.gov/bill/117th-congress/senate-bill/1265
14. V. Zinchenko, “Sustainable development goals as an integrative basis of the global public strategy for the effectiveness of ecology, education and science at all levels”, in *IOP Science: Earth and Environmental Science* **635,** 012012 (2021).
15. V. Zinchenko, M. Boichenko and O. Polishchuk. “Education for Sustainable Development and Beyond It”, in *Proceedings of the VIII International Scientific and Practical Conference 'Current problems of social and labour relations'* **527** (2021), pp. 751-755.
16. V. Zinchenko, M. Boichenko and O. Polishchuket, “Critical thinking education as a prevention of AI riskis and basis of AI and Open Science based Sustainable development”, in: *The 8th Technium international Conference «Sustainable Future and Technology Development»* **8** (2021), https://techniumscience.com/
17. S. Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (Profile Books, London, 2019).
18. S. Zuboff, N. Moellers, D. M. Wood, and D. Lyon, Surveillance & Society **17(1/2)**, 257-266 (2019).