Edukacja Filozoficzna 79/2025
ISSN 0860-3839
DOI: 10.14394/edufil.2025.0002
ORCID: 0000-0002-5339-781X

Ideological Limits to Ethical Artificial Intelligence

Luka Perušić
(University of Zagreb, Faculty of Humanities and Social Sciences)

Abstract: The paper analyzes the current use of *ethical* artificial intelligence (AI), argues that there are ideological limits to it, and discusses these limits. The topic is of particular relevance to research on the social implementation of AI systems, as ideological underpinnings are not easy to identify and ideology research is underrepresented in research on AI phenomena. The first section analyzes what counts as ethical in ethical AI systems. The second section classifies the dimensions of the ethical in AI systems, highlights their interrelationships and applies *forness* as a key concept that helps narrow the focus on the ideological component of ethical AI. The third section describes the presence of ideology in ethical AI and clarifies the limits it imposes on AI as a general phenomenon that undoubtedly has the potential to contribute to a more humane society, but is severely constrained by ideology.

Key words: morality, ethic, ethics, politics, ideology, forness, artificial intelligence, artificial agency

1. Understanding the "Ethical" in Ethical Artificial Intelligence[1]

1.1. The Status of the Ethical

In recent years most of the economically leading countries, supranational entities, and international expert organizations, together with the most influential technology companies, are striving to create workable frameworks for the development, implementation, use and evaluation of artificial intelligence (AI) systems. The level at which AI has suddenly been taken seriously is technologically unmatched, with the European Union (EU) at the forefront in terms of intensity, scope and thoroughness, culminating in the proposal of the Artificial Intelligence

[1] I sincerely thank the reviewers and editors for their efforts to advance the quality of the paper.

Act.[2] In the spectacle of initiatives, guidelines, strategies, policies and legislative preparations to harness the advances in AI development and implementation, the question of ethics has been ever present, and the notion of ethical AI could neither be avoided nor evaded. Scholars working in the fields of morality, ethics and AI, as well as policy makers and jurists dealing with ethical AI, have approached these problems with different emphases:

> [Strategies for approaching (ethical) AI] range from regulations, law, codes of conduct, attempts to design AI with safety and ethics uppermost, attempts to build ethics into design process, specific strategies such as attempts to understand and mitigate bias, and so on. Some focus on current issues; some focus on longer-term and more speculative questions, such as possible dangers of superintelligence. Some issues are concrete and specific; some are more general, wide-ranging, or foundational. Some approaches lean towards the view that AI presents a threat that we might lose control of ourselves and of our values and that we need radical shifts to deal with the world that is coming. Other approaches are more sanguine and diligently tread the path of trying to ensure that the technologies that are being developed and used fit within current frameworks of value in approaches broadly labelled "value alignment."[3]

However, the discourse on being "ethical" continues to most commonly perpetuate the idea that *ethical*[4] refers to having a set of principles that instruct on proper conduct towards others or on what values should be embodied and manifested. Although the approach may seem functional, in the contemporary technological forefront society *the ethical* was never made fundamental neither in terms of nurturing and education nor in terms of legislation – in the context of the triple helix complex (military, industry, academia), it was systematically relegated to "playing the role of a bicycle brake on an international airplane"[5]

[2] European Parliament, P9_TA(2023)0236: Artificial Intelligence Act, June 2023, amendments; cf. Council of the European Union, *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: Analysis of the Final Compromise Text with a View to Agreement*, no. Cion doc. 8115/21, Brussels, 26 January 2024.

[3] P. Boddington, *AI Ethics: A Textbook*, Springer, Singapore 2023, p. 6.

[4] Almost always derived from the word *ethics*, even though it should derive from *ethic*, as the latter is a set of action-guiding principles related to moral behaviour, while the former is a branch of philosophy.

[5] U. Beck, *Gegengifte. Die organisierte Unverantwortlichkeit*, Suhrkamp Verlag, Frankfurt am Main 1988, p. 194.

already in the 1980s. Contemporary studies support this argument by showing that ethical norms have *near-zero* influence on the tech community, both the student population and working experts.[6] This raises the possibility that incentives from major players to build an ethical AI and use AI ethically may not be ethical and may not support the consistency between moral behaviour and legislation. A brief insight into the motives behind the formation of ethically aligned products was given in 2016 by a Mercedes representative, Christoph von Hugo, whose comment was one of the first public comments on the moral issues related to self-driving cars made by companies producing such vehicles. Von Hugo stated that self-driving Mercedes cars "would always prioritize their owners," before changing his statement after a public outcry.[7] It is a preference that is understandable from the perspective of a product seller, but not from the perspective of the fundamental rules and laws of traffic regulation or from the perspective of non-driving members of the contemporary social environment. The relegation of the ethical to an inferior position produced at least two consequences: (1) the possibility to manipulate the notion of the ethical for the protection of personal gain against the other, and (2) the relativization of morality.

In relation to the first consequence, the devaluation of the ethical has reached a new level, especially in the context of climate change and sustainability, with technology companies themselves expressing ethical concerns for twofold effect: (1) *legislative*, because by pretending to deal with the ethics of their own inventions they are stalling talks about the regulation of their activities and products,[8] and (2) *commercial*, because by expressing their ethical standpoints and pasting them on the "cover" of their brand they market themselves as trustworthy humanists contributing to the creation of a better society, and thus a better label for stakeholders to spend money on.[9] The epitome of this misleading strategy is their

[6] L. Munn, *The Uselessness of AI Ethics*, "AI and Ethics" 2023, Vol. 3, No. 3, p. 872, https://doi.org/10.1007/s43681-022-00209-w; T. Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines*, "Minds and Machines" 2020, Vol. 30, No. 1, p. 108, https://doi.org/10.1007/s11023-020-09517-8. Both papers link to several different studies supporting the argument empirically.

[7] S. Nyholm, *The Ethics of Crashes with Self-Driving Cars: A Roadmap, I*, "Philosophy Compass" 2018, Vol. 13, No. 7, e12507, p. 5, https://doi.org/10.1111/phc3.12507.

[8] L. Munn, *The Uselessness of AI Ethics*, op. cit., p. 872.

[9] N. de Marcellis-Warin et al., *Artificial Intelligence and Consumer Manipulations: From Consumer's Counter Algorithms to Firm's Self-Regulation Tools*, "AI and Ethics" 2022, Vol. 2, No. 2, p. 264, https://doi.org/10.1007/s43681-022-00149-5.

private, unregulated development of AI solutions, veiled by their publicly expressed concern about the apocalyptic coming of artificial general intelligence – a self-aware, adapting, and learning autonomous AI that may take human beings "out of the picture." The practice is as absurd as publicly warning that all of human civilization could die from a deadly virus while privately developing it in the lab; however, misleading and entertaining the public serve the purpose of allowing the companies the freedom to develop AI systems for these companies' gain.

In addition, Thilo Hagendorff highlights that "ethics can also simply serve the purpose of calming critical voices from the public, while simultaneously the criticized practices are maintained within the organization."[10] Deconstructed, the practice is a form of "ethics washing," a sibling to the well-known phenomenon of greenwashing. Ethics washing is "the practice of visibly, sometimes ostentatiously, showing to the world that one is taking great care to attend to ethics, while in reality, doing little or nothing."[11] A notorious example of such behaviour is the inappropriate firing of Timnit Gebru by Google after Gebru insisted on publishing a report that demonstrated how AI systems could generate racial results, while simultaneously presenting the company as a leader in ethical standards.[12] Granted, it would be unconvincing to claim that ethics washing – and the entirety of ethical devaluation processes – apply to the entirety of the AI systems development landscape. AI development spans from developing systems that assist in mountain rescue missions, through news feeds, to unmanned ground vehicles for assault combat, and there are certainly many authors who actively engage in discussions of the best possible utilization of AI systems. For example, Seng W. Loke, who, analyzing the game theory problem in the context of interaction among autonomous AI systems, has proposed the prime rule "Cooperate first" as "a good candidate for a universalizable maxim (i.e. 'seeking first to cooperate' could be willed as a strategy for everyone)" that would possibly manage the autonomous interaction of AI systems in a vehicle network for the benefit of all participants.[13] However, here I focus on examples generated by

[10] T. Hagendorff, *The Ethics of AI Ethics*, op. cit., p. 100.

[11] P. Boddington, *AI Ethics*, op. cit., p. 21; cf. L. Munn, *The Uselessness of AI Ethics*, op. cit., p. 872.

[12] For a good overview of the case and the understanding of multiple layers of misconduct involved in the process of firing Gebru, see T. Simonite, *What Really Happened When Google Ousted Timnit Gebru*, Wired, 8.06.2021, URL: https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/.

[13] S.W. Loke, *Designed to Cooperate: A Kant-Inspired Ethic of Machine-to-Machine Cooperation*, "AI and Ethics" 2022, Vol. 3, No. 3, p. 992, https://doi.org/10.1007/s43681-022-00238-5.

powerful global entities that might not be as clear as they initially appear. For example, although Microsoft as a global tech company is one of the parties that have committed to apply UNESCO's *Recommendation on the Ethics of Artificial Intelligence* in 2022,[14] it has invested over $13 billion into OpenAI, which has already been shown to favour exploitative practices.[15]

A second consequence of the relegation of the ethical to an inferior position is the multiplication of various ethical codes that are constantly proposed regardless of other efforts, resulting in a plethora of ethical proposals that are not supported by legal systems in terms of sanctions. Combined with the evidence of cultural differentiation in the world, a relativistic image of ethics emerges and a negative view of ethics as arbitrary or limited. Given the case, some authors argue that it is pointless to discuss the ethics of AI (and thus ethical AI) and that we should focus on law-abidingness and accountability.[16] Roman V. Yampolskiy raises the old but standing problem of *legal positivism* or *legal blindness*, in the sense that what is allowed or forbidden by law may be unethical (for instance, ban on same-sex marriage and acceptance of underage marriage), and later cannot be prosecuted because it was acceptable from the perspective of the law that was in effect at the time. In that regard, the EU's Artificial Intelligence Act is a peculiar case which serves well to clarify what "ethical" stands for in the phrase "ethical AI."

1.2. The Meaning of the Phrase "Ethical AI"

The European Commission (EC) has accepted a document titled *Ethics Guidelines for Trustworthy AI*, which considers "ethical AI" to be an AI system following the set of principles labelled by the Commission as "ethical" (respect for human autonomy, prevention of harm, fairness, and explicability).[17] This selection

[14] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, Paris 2022.

[15] More on this in the next section. For the information on Microsoft, see J. Liboreiro, *European Regulators Put Microsoft's $13 Billion Bet on OpenAI under Closer Scrutiny*, EuroNews, 9.01.2024, URL: https://www.euronews.com/my-europe/2024/01/09/european-regulators-put-microsofts-13-billion-bet-on-openai-under-closer-scrutiny.

[16] For example, computer engineer Roman V. Yampolskiy, who stated this before the political world took AI seriously. See R.V. Yampolskiy, *Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach*, in: *Philosophy and Theory of Artificial Intelligence*, ed. V.C. Müller, Springer-Verlag, Berlin 2013, pp. 389–390.

[17] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, European Commission, European Union 2019, pp. 12–13. This document is cited in later documents on AI, including the documents related to the Artificial Intelligence Act proposal that

was drawn from the EU Charter of Fundamental Rights and later expanded in the finalization of the Artificial Intelligence Act proposal. Ben Wagner explains that "EU fundamental rights are not understood as fundamental rights but rather as ethical imperatives to be complied with in a non-binding fashion."[18] In fact:

> In this sense these are "potential fundamental rights," developed under the shadow of hierarchy of the European Commission. They certainly cannot be claimed at present and if these potential fundamental rights are "violated" (whatever that means in the context of ethical commitments to uphold fundamental rights) they would be no legal recourse of any kind available. Indeed, it is in fact likely that these rights would actively need to be violated frequently and these violations would need to be made public widely, in order for the European Commission to be willing to do anything about their actual violation.[19]

However, to define "ethical AI", the EC created a single concept – *trustworthy AI* – composed of three distinct phenomena – law (the AI has to be law-abiding), ethics (the AI has to follow a set of action-guiding principles), and technics (the AI has to be robust). By doing so, the EC's proposal merged law and ethics into a single entity, even though it itself differentiates between law and ethics like Yampolskiy does, by strongly focusing on AI system *solution* via *value alignment* and in that way, at least on the surface, further attempted to prepare ground for subduing the environment which creates AI systems and the actor network that uses AI systems, doing so outwardly, that is, making the solution itself the starting point, thus going beyond "regulation by design."[20] For example, when the proposal states that an AI system should be "transparent," it means that all human and non-human elements in its entire life cycle have to align to the value of traceability and explainability[21] for it to successfully retain the accepted property, and so practices such

was formally adopted in June 2023. For a review, see L.A. DiMatteo, *Artificial Intelligence: The Promise of Disruption*, in: *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, eds. L.A. DiMatteo, C. Poncibò, M. Cannarsa, Cambridge University Press, Cambridge 2022, pp. 12–14.

[18] B. Wagner, *Ethics as an Escape from Regulation: From "Ethics Washing" to Ethics-Shopping?*, in: *Being Profiled: Cogitas Ergo Sum. 10 Years of Profiling the European Citizen*, eds. I.E. Bayamlıoğlu et al., Amsterdam University Press, Amsterdam 2018, p. 85.

[19] Ibid.

[20] L.A. DiMatteo, *Artificial Intelligence*, op. cit., p. 14.

[21] European Parliament, P9_TA(2023)0236: Artificial Intelligence Act, Amendment 213, Article 4 a (p. 127).

as psychological targeting,[22] fake news, hate generation, preference recognition,[23] etc., should be considered for prohibition. If this conception was to be enforced with strict regulation and the appropriate administrative support, it might be representative of how potentially dangerous and exploitative new technologies could be used where they benefit humanity by aiming to create systems based on technical invention. It could be understood as a way of addressing the general problem of technical inventions taking control of social processes.

The idea of calling AI systems "ethical" further stems from the development of AI systems that exhibit autonomous behaviour, thus resembling a subject. This is highly debatable because most of what is considered "autonomous" in discussions on AI is most likely a more complex form of automation. In the simplest terms, a system that was automatized is a system that will once initiated continuously carry out the specified task by itself until completed without deviation –

[22] "Recent research in the field of computational social sciences […] suggests that people's psychological profiles can be accurately predicted from the digital footprints they leave with every step they take online. For example, people's personality profiles have been predicted from personal websites, blogs, Twitter messages, Facebook profiles, and Instagram pictures. This form of psychological assessment from digital footprints makes it paramount to establish the extent to which behaviours of large groups of people can be influenced through the application of psychological mass persuasion – both in their own interest (e.g., by persuading them to eat healthier) and against their best interest (e.g., by persuading them to gamble)" – S.C. Matz et al., *Psychological Targeting as an Effective Approach to Digital Mass Persuasion*, "PNAS" 2017, Vol. 114, No. 48, p. 12714, https://doi.org/10.1073/pnas.1710966114.

[23] An extreme case of preference recognition is AI's ability to detect sexual orientation solely by observing facial images and with much higher accuracy than human beings. These particular results were published by Yilun Wang and Michal Kosinski (Stanford University) in 2018, in a paper titled *Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images*: "Their decision to do the study at all, despite the evident risk to people living in countries where homosexuality is illegal, is justified by the authors in terms of the fact that if it is possible, then it represents a risk and should be public" – A. Campolo, K. Crawford, *Enchanted Determinism: Power without Responsibility in Artificial Intelligence*, "Engaging Science, Technology, and Society" 2020, Vol. 6, p. 12. Cases related to facial recognition are especially troublesome because they are notorious for the lack of certain interpretability of how and why the results are generated. See L.D. Introna, D. Wood, *Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems*, "Surveillance and Society" 2004, Vol. 2, Nos. 2–3, pp. 177–198, especially pp. 183–184. Developers are struggling to this day to reduce the black-box effect. For example, Wang and Kosinski also did not know exactly why their AI system is able to detect sexual orientation, and this is also becoming the problem in analysis or understanding "whether each action is performed in a responsible or ethical manner" – I. Gabriel, *Artificial Intelligence, Values, and Alignment*, "Minds and Machines" 2020, Vol. 30, No. 3, p. 412, https://doi.org/10.1007/s11023-020-09539-2.

"the machine is on; it runs its course."[24] To differentiate from simple automated systems,[25] it can be said that an *autonomous system* is "a system situated in an environment that senses the environment and acts on it in pursuit of its own agenda, in such a way that its actions can influence what it later senses."[26] Moreover, the capabilities of "learning," "adaption," and "choice-making" are added to such systems, with some authors emphasizing that it is about objects having "unsupervised activity."[27] But all these notions, which we would usually apply to living beings – perception, learning, adaption, having an agenda, choice-making, etc. – do not really transfer to machines. They are an *artificial* resemblance of organic capabilities because they are neither equivalent to capabilities found in living organisms nor the way in which they manifest can be found in living organisms. The unnecessary humanization of machines is maybe best seen in the use of the notion of "own agenda" instead of "specified task defined by the external user." So when encountered in the discourse, phrases signifying human behaviour and capabilities should be thought of as technical terms derived from the original notion applicable to living beings because of their orientational value in the knowledge landscape. Likewise, "autonomous" could be understood as higher-order automation because there is nothing in autonomous AI processes that differs from the fundamental trait of being a system that is continuously carrying out specified tasks by itself until completed without deviation. For this reason they can be only thought of as implicit subjects and their "morality" is only *functional* at their best, "where the machines themselves have the capacity for assessing and responding to moral challenges,"[28] but they retain moral inacces-

[24] H.M. Roff, *Artificial Intelligence: Power to the People*, "Ethics and International Affairs" 2019, Vol. 33, No. 2, p. 128, https://doi.org/10.1017/S0892679419000121.

[25] A class of auto-initialization lower than automation is automatization. "Automatic systems, such as a toaster in the civilian world or, to use a military example, an explosive triggered by a tripwire, respond mechanistically to environmental inputs. Automated systems, by contrast, operate based on multiple pre-programmed logic steps" – M.C. Horowitz, *Artificial Intelligence, International Competition, and the Balance of Power*, "Texas National Security Review" 2018, Vol. 1, No. 3, p. 40.

[26] S. Franklin, *History, Motivations, and Core Themes*, in: *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish, W.M. Ramsey, Cambridge University Press, Cambridge 2014, p. 27. Cf. H.M. Roff, *Artificial Intelligence*, op. cit., pp. 129–130.

[27] C. Allen, W. Wallach, *Moral Machines: Contradiction in Terms or Abdication of Human Responsibility?*, in: *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. P. Lin, K. Abney, G.A. Bekey, The MIT Press, Cambridge, MA, 2012, p. 55. "Unsupervised" to, still, "execute tasks on the designer's behalf" – E. Alonso, *Actions and Agents*, in: *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish, W.M. Ramsey, Cambridge University Press, Cambridge 2014, p. 235.

[28] Ibid., p. 57.

sibility – they cannot know that their operations are "moral," and what is or is not a "moral challenge" is recognized by human beings, not autonomous AI systems.

"Ethical AI" is altogether a clumsy expression because it subsumes the multitude of meanings hidden under the abbreviation "AI" and perpetuates the modern trend of the technical connectivity of moral subjectivity to non-living, non-self-conscious objects via norms.[29] "Ethically aligned AI", as proposed by IEEE Global Initiative,[30] is a better expression because it tells us that AI was aligned by something to mediate conduct towards itself and others without itself being a moral subject. The expression "ethical AI", although not to my scientific liking, is, however, pragmatic and applied widely. It should first be understood in the broadest sense as an AI system that, by its very existence, embodies preferred principles related to optimal moral behaviour in the human sense. Ethical AI is thus an AI system whose construction and performance is subject to predefined norms and values that are considered socially acceptable. However, what is "socially acceptable" in its universality is challenged by the realism of cultural relativism and personal preferences. "Ethical AI" as a term hides its structural complexity essentially related to the *ideological* component by which the social acceptability inherent to the term is limited.

In this paper, *ideology* is understood as "systematized ideas that, if followed in a prescribed manner, will lead to a preferred social outcome."[31] The preference of social outcome may aim at its possible universality, but it may not. In an armed conflict between two states, nations, ethnic groups, tribes, etc., social preferences are clashed despite some of them being possibly compatible. The existence of different cultural set-ups that generate different social preferences, for example, the acceptability of death penalty for apostasy, tells us that there are only two ways of developing and applying ethically aligned AI, either with the aim of supporting

[29] In philosophy this is usual for American and Dutch new waves of the philosophy of technology, and Luciano Floridi's circle of influence.

[30] IEEE Global Initiative, *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*, 2019, URL: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf.

[31] N. Chitty, S. Dias, *Artificial Intelligence, Soft Power and Social Transformation*, "Journal of Content, Community and Communication" 2017, Vol. 6, No. 3, p. 1. Of course, there is a plethora of slightly different perspectives and uses of the concept of ideology, and the use of a more prominent approach, such as that of Karl Mannheim, Karl Marx, Marxists, Karl Jaspers, Herbert Marcuse, Jürgen Habermas, David Bloor or Michel Foucault, would certainly be useful for the analysis. However, Chitty and Dias's formulation is very effective for the discourse on ethical AI, especially since of those who refer to the concept of ideology at all in their work on AI, the majority of authors who have mentioned it use it without meaningful relevance to the general research on AI.

universally acceptable social preferences or otherwise. The latter is usually a sign of ideology building the foundation for a particular action. What this paper proceeds to show is that "ethically aligned AI," albeit discussed as if it is a matter of universal moral code, in practice embodies systematized ideas for a preferred non-universal outcome that is presented as an ethically aligned product. "Ideological limits" emanate from the core system of ideas embodied in the product or its application, in that any "ethical alignment" – either as engineered or applied – becomes a set of non-universal preferences that benefit some, but not all. In that sense, "ethical" becomes a simple descriptive term for *having a set of dispositional principles for expected conduct*, not a term referring to what is truly right or wrong, good or evil, morally permissible or impermissible, that we may further find to be universal or universalizable. The "ideological limit" thus denotes a boundary beyond which "ethical" is just a preference construct for a restricted gain.

The following section categorizes the dimensions of "ethical AI" and discusses the details of difference among them. The classification serves to show different ways of how the supposed ethical alignment can be carried out and how it relates to a difference between engineering practice, legal compliance, and social acceptability, for the purpose of showing how various instances of the problem of ideology come to the fore. These instances are then exemplified and discussed in the third section.

2. Classifying the Ethical in Artificial Intelligence

2.1. Basic Distinctions

To successfully tunnel through the ethical AI systems problem network, the simplest approach is to separate the presumed content into fundamental categories:

- ethical design of AI systems
- ethical development of AI systems
- ethical behaviour of AI systems
    - non-autonomous
    - autonomous
    - self-aware non-autonomous
    - self-aware autonomous
- ethical use of AI systems.

An AI system can be designed so that, for example, all information and actions related to the activity of the AI system are transparent and accessible/readable by anyone who has elementary information and digital literacy skills. This does not mean that the high-value data was properly tested when it was developed, and if it was, it does not mean that it was obtained in a fair way or without exploiting intellectual property loopholes. We can have an ethical AI compliant with current legal systems that appears socially acceptable, but was developed unethically. Even if the data was properly tested and obtained in a fair way, it still does not mean that the AI system was trained or managed ethically. One paradigmatic example is the functioning of OpenAI, a company that developed a sensible, amusing and reasonably useful application, ChatGPT, by using cheap labour,[32] switching from non-profit organization to profit company after achieving its developmental goal on the basis of donations,[33] and exploiting the uncontrolled data flow of the entire accessible Internet, including collective non-profit common-good efforts such as Wikipedia, to build its database for "training" AI for a service that then became privileged and now consumes 500 millilitres of water per 5 to 50 queries and spends an energy equivalent of up to 33,000 households per day.[34] The product may appear socially acceptable, it may offer clean and valuable data, but its developers may have exploited legal loopholes and weak links in the social environment for the product to become possible and feasible. The case is akin to enjoying an Apple smartphone that contains cobalt obtained through child labour in Congo mines.

[32] B. Perrigo, *OpenAI Used Kenyan Workers on Less Than $2 per Hour to Make ChatGPT Less Toxic*, "Time," 18.01.2023, URL: https://time.com/6247678/openai-chatgpt-kenya-workers/.

[33] C. Nduka, *How OpenAI Transitioned from a Nonprofit to a $29B For-Profit Company*, Hackernoon, 28.03.2023, URL: https://hackernoon.com/how-openai-transitioned-from-a-nonprofit-to-a-$29b-for-profit-company.

[34] Water consumption estimates were pre-reported in C. Novo, *The Water Cost of Artificial Intelligence Technology*, "SmartWaterMagazine," 12.09.2023, URL: https://smartwatermagazine.com/news/smart-water-magazine/water-cost-artificial-intelligence-technology. For a broader survey on AI's background water footprint, see the paper the report is based on: P. Li et al., *Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models*, arXiv:2304.03271 [cs.LG], https://doi.org/10.48550/arXiv.2304.03271. Energy estimates were a result of Sajjad Moazeni's research; basic information can be found in S. McQuate, *UW Researcher Discusses Just How Much Energy ChatGPT Uses*, University of Washington, 27.07.2023, URL: https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/. For an unrelated study on the growing energy footprint of AI, see A. de Vries, *The Growing Energy Footprint of Artificial Intelligence*, "Joule" 2023, Vol. 7, No. 10, pp. 2191–2194, https://doi.org/10.1016/j.joule.2023.09.004.

Similarly, an AI system may be both designed and developed in accordance with the expected conduct, that is, "ethically," but depending on what the AI system actually is, how well it is designed and developed, and how its use is regulated and limited, its ethical design and development may be denied in practice. A non-autonomous AI system, for instance, dialogic software such as ChatGPT, may provide dangerously inaccurate information about, for example, human conflict history or social status, regardless of the developer's best possible intentions, and could offer wrongful guidance in conduct to those who might ask for such a thing.[35] An autonomous AI system, such as the one implemented in an armoured combat vehicle, can be damaged, hacked or corrupted during war. The result can be an environmental miscalculation resulting in underage civilian casualties through action independent of human guidance. AI systems applied in predictive policing have already showed disastrous results because they are biased, racial, suggest oppressive monitoring practices and hamper elementary human rights.[36] Self-aware autonomous AI systems, which are currently only speculated about, have the same potential range of possible ethical misconduct as humans.

Ultimately, if an AI system were designed and developed in complete compliance with expected ethics and "behaved" accordingly, it could still be misused and exploited for unethical purposes. Unethical use must not be conflated with ethical AI, but the distinction still has to be made. For example, an AI system can be developed to simply track, record and analyze the movements of life systems. Such a system could be used to track animal populations in an ecosystem to help preserve biodiversity. But it can also be used to track undesirables, as in the two high-profile African cases where the Chinese company Huawei assisted the Ugandan and Zambian governments in tracking political opponents by sell-

[35] For example, in March 2023, the Belgian daily newspaper *La Libre* reported that a man had allegedly committed suicide after continuously exchanging information with an AI chatbot on an app called Chai. The man had previously been "increasingly pessimistic about the effects of global warming" and had isolated himself from family and friends in the pursuit of understanding the problem through the use of the dialogical AI system. See C. Xiang, *"He Would Still Be Here": Man Dies by Suicide after Talking with AI Chatbot, Widow Says*, Vice, 30.03.2023, URL: https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says.

[36] For an overview, see Fair Trials, *Automating Injustice: The Use of Artificial Intelligence and Automated Decision-Making Systems in Criminal Justice in Europe*, 9.09.2021, URL: https://www.fairtrials.org/articles/publications/automating-injustice/.

ing them AI-based equipment.[37] Here, too, is where ideological limits to ethical AI can be considered. From the perspective of EU citizens these practices may be considered socially unacceptable and legally questionable. Yet they did become a social reality for Uganda and Zambia, with whom we are connected at least through accepting Huawei products in our local stores and buying them for our business and amusement, and the legal system in Uganda and Zambia can support that kind of technological use. From the perspective of the upholder of the current state of affairs, neither the AI systems are unethical nor their use is unethical because fighting against the government is viewed as unethical. In addition, in the case of self-aware AI systems, even if everything to do with design, development, behaviour and use is ethically formidable, the instrumentalization of a self-aware entity is at least morally questionable, especially if such a system begins to pursue on its own an end that deviates from the intended means.

2.2. The Forness of Artificial Intelligence Systems

The stratification of ethical phenomena related to AI systems stems from the nature of AI systems as made things. Firstly, AI, narrowly understood as a study field of computer science and engineering,[38] broadly being a "wide range of technologies or an abstract large-scale phenomenon,"[39] is essentially an imitative solution[40] that becomes implemented into machine systems performing actions

[37] J. Parkinson, N. Bariyo, J. Chin, *Huawei Technicians Helped African Governments Spy on Political Opponents*, "Wall Street Journal," 15.08.2019, URL: https://www.wsj.com/articles/huawei-technicians-helped-african-governments-spy-on-political-opponents-11565793017#comments_sector. Uganda and Zambia belong to the top third of most corrupt countries in the world, as established by Corruption Perceptions Index.

[38] S. Franklin, *History, Motivations, and Core Themes*, op. cit., p. 15; S.M. Liao, *A Short Introduction to the Ethics of Artificial Intelligence*, in: *Ethics of Artificial Intelligence*, ed. S.M. Liao, Oxford University Press, Oxford 2020, p. 3. A variant definition to AI as discipline was given by Iason Gabriel as "the design of artificial agents that perceive their environment and make decisions to maximise the chances of achieving a goal" – I. Gabriel, *Artificial Intelligence, Values, and Alignment*, op. cit., p. 412.

[39] T. Hagendorff, *The Ethics of AI Ethics*, op. cit., p. 111.

[40] Authors discussing this variously refer to mimicry, imitation, and simulation. These are not entirely precise terms, but they are applicable to different contexts. Mimicry can certainly be applied to end products that have been biomorphized to be more accessible and user-friendly, or to actions that exhibit behaviour derived from the function of mimicry. AI systems certainly simulate in the broader sense of the word (originally, simulation referred to the creation of running models for the purpose of predicting its outcomes or representing it), but nevertheless they are based not only on trying to duplicate behaviour and outcomes, but also on duplicating

that use the distinct computation method resembling thought-processing, and appears as a material (physical/digital) entity performing tasks translated into understandable output through the computer interface and hardware shell.[41] It is a process that "exploits" a "realisation that nature, or human nature, works a certain way,"[42] constructed into material systems that combine multiple natural "effects" into a "chain of effects"[43] to our expected working advantage. Because AI was invented, designed, developed, and deployed by human beings, it should be understood as a *product* – a non-living produce of human beings. Being a software in a hardware shell, as products AI systems have all the characteristics of constructed artefacts – "object made by a human being that is not naturally present but occurs as a result of the preparative or investigative procedure by human beings."[44] For such an object to be, matter is "transformed such that the resulting physical construction has certain capacities or shows a particular kind of behaviour,"[45] attaining the status of objects that have a specific "practical 'for-ness,'"[46] which is generated by human activity. The element of *forness* explicates the fundamental aspect of artefacts as human-made *conveyants*. Being "for something" means that there is an interactor that will activate properties of conveyance in a specific artefact and cause an effect manifesting within the artefact and to its environment. For that reason, scholars in the second half of

certain internal processes or abilities of living beings, which would also make them emulative systems. However, all three concepts serve the purpose of imitation for a specific purpose. The aspect of imitation is important for understanding AI in relation to the general forms of machine learning today. Although most AI systems today use machine learning, imitation can also be achieved through other means, such as programmed execution rules masquerading as intelligent behaviour, as was the case with so-called expert systems in the 1980s. Machine learning by itself is "creation of software-based algorithms that build a mathematical model based on data, that can make decisions, predictions or perform tasks without being specifically programmed to do these tasks," usually attributed to AI (H. Seaton, *The Construction Technology Handbook*, John Wiley & Sons, Hoboken 2021, p. 102). This means that AI is an abstract idea, currently based only on machine learning, but it need not be so. It can be seen as a deployable capability to imitate for a specific purpose. The more complex the solution (expert system vs humanoid robot for elderly care), the clearer the attribute of imitation.

[41] We should not rule out the possibility that AI systems in the future will not be computer based, which will certainly further blur the line between artificial and natural agency.

[42] H. Seaton, *The Construction Technology Handbook*, op. cit., pp. 2–3.

[43] Ibid., p. 4.

[44] P.E. Ekmekci, B. Arda, *Artificial Intelligence and Bioethics*, Springer Nature Switzerland, Cham 2020, p. 17.

[45] P. Kroes, *Technical Artefacts: Creations of Mind and Matter*, Springer, Dordrecht 2012, p. 3.

[46] Ibid., p. 4.

the 20th century slowly began to conceptualize human products – technical artefacts foremostly – as mediators.

> When a technological artefact is used, it facilitates people's involvement with reality, and in doing so it coshapes how humans can be present in their world and their world for them. In this sense, things-in-use can be understood as mediators of human-world relationships. Technological artefacts are not neutral intermediaries but actively coshape people's being in the world: their perceptions and actions, experience and existence.[47]

As such, they may "ascribe new value to human beings, nonhuman things, and even to 'non-things' like future people and animals."[48] An AI system must be understood as a mediating technical product so that we can observe how its manifestation passes through phases of operational dimensions and comes into contact with the human lifeworld in which people articulate their preferred environment, for example, warfare and the promotion of political exceptionalism versus peace mediation and cosmopolitanism. This mediation of value grants them "normative power"; they are "examples of how code is law as well as how code creates law, or rather produces norms,"[49] which can be demonstrated by any number of applications, from the norm the AI imposes in filtering out discussions on social networks, through influence in medical or legal analysis and choice-making, to the selection of feasible workers, mortgages, or the creation of "new rules of interaction between economic agents" to "create a new form of

[47] P.-P. Verbeek, *Moralizing Technology: Understanding and Designing the Morality of Things*, The University of Chicago Press, Chicago 2011, pp. 7–8.

[48] L. Magnani, *Morality in a Technological World: Knowledge as Duty*, Cambridge University Press, Cambridge 2007, p. 13. Magnani gives an example: "Think for a moment of cities with extensive, technologically advanced library systems in which books are safely housed and carefully maintained. In these same cities, however, are thousands of homeless human beings with neither shelter nor basic health care. Thinking about how we value the contents of our libraries can help us to reexamine how we treat the inhabitants of our cities, and in this way, the simple book can serve as a moral mediator."

[49] G. de Gregorio, *The Normative Power of Artificial Intelligence*, "Indiana Journal of Global Legal Studies" 2023, Vol. 55, p. 3, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4436287. Cf. L.A. DiMatteo, *Artificial Intelligence: The Promise*, op. cit., p. 11: "Lawrence Lessig has argued that coders and software programmers, by making a choice about the working and structure of IT networks and the applications that run on them, create the rules under which the systems are governed. The coders therefore act as quasi-legislators. In other words, 'code is law' is a form of private sector regulation whereby technology is used to enforce the governing rules."

social order."[50] The ability to detect patterns or specifics unavailable to human beings carries the capacity for formulating norms because the computational result widens the perspective on reality. When viewed in the light of the Artificial Intelligence Act, AI systems are basically used as enhancers to the preferred norms and generators of new incentives. This phenomenon will become even more evident when (if) there will be an artificial general intelligence, as the object will begin to constitute norms for itself. Moreover, as products in the commercial sense and as tools of commercialization, AI systems acquire an additional dimension of use and mediation that must be taken into account, especially since AI became a symbol of the ongoing Fourth Industrial Revolution as the first such revolution originating from the private sector.[51] The market is the only playing field of the private sector, and "the profit motive ultimately drives markets."[52]

For example, a non-autonomous, non-self-aware AI system can be designed, developed and used in medicine in a completely ethical way to reduce tremor in Parkinson's disease patients, but what is the cost of restoring the person's quality of life? Is such conditioned use of AI ethical? Furthermore, an AI system may be entirely ethically designed and developed and used to monitor geographic movements for positive purposes, but a company may decide to capitalize on its product by selling it to parties who use it with the intention of harming people, regardless of the terms of trade. Furthermore, consider the imbalance between government and the population: data-collection technologies and AI systems used in the public sector, especially in government institutions, increase knowledge about the population and the ability to exercise power over the population, while the population knows less and less about the government's activities and is not granted any privileges to make its plans and activities a matter exempt from the law and kept secret from the public (for example, military operations, international negotiations, surveillance of public space, etc.). Benjamin Baez, following Grigori Perelman, puts it aptly:

[50] Å. Melkevik, *The Internal Morality of Markets and Artificial Intelligence*, "AI and Ethics" 2023, Vol. 3, No. 1, p. 115, https://doi.org/10.1007/s43681-022-00151-x.

[51] To the point where China has decided to change its usual state politics, and support non-state, private companies in developing AI. See W.A. Carter, W.D. Crumpler, *Smart Money on Chinese Advances in AI: A Report of the CSIS Technology Policy Program*, Center for Strategic and International Studies, Washington 2019, p. 5.

[52] N. de Marcellis-Warin et al., *Artificial Intelligence and Consumer Manipulations*, op. cit., p. 261.

> [A]long with nation-states, large corporations enjoy great control over information "resources" (which include actual workers in the information economy, such as systems analysts, academics, etc.), and combined with the fact that these large corporations own formerly public resources because of privatization, and that media is increasingly becoming concentrated in these corporations, we certainly can say without qualification that the increasing centralization and monopolization of information is not overstating matters. What this means, as Perelman points out, is that in addition to withholding information from the public, the owners can also manipulate and censor information, distorting the public's understanding of situations, and making it more difficult for people to challenge what is happening to them (Perelman, 1998, p. 78).[53]

From AI being viewed as a mediating technical product, we can derive its nature, on the one hand, as a *weapon*, and on the other, as an *artificial agent*.[54]

Being a mediating technical product, an AI system is always a product for something, a tool. When used for offence or defence, it must be considered a weapon, not so much because of the possibility of incorporating it into other weapons, but rather because the wide range of AI systems can be weaponized. Weaponization of AI systems should not be understood narrowly in the sense that an AI system designed as a mere tool is converted exclusively into a weapon. AI systems can serve as a weapon and be a tool at the same time. For example, police personnel using AI surveillance systems may target or monitor specific groups to gain personal benefit, although the use is certainly monitored and restricted to some degree. In the context in which a tool is used offensively against a living being, it behaves like a weapon, whether or not this was intended and whether or not there is a direct physical interaction typical of classical weapons, since the end goal is to endanger life.

The higher order of AI system utilization is the deployment of artificial agents because, in addition to computing advantages that imitate reasoning, an AI sys-

[53] B. Baez, *Technologies of Government: Politics and Power in the "Information Age"*, Information Age Publishing, Charlotte 2014.

[54] Authors would usually use words such as *actor*, *agent*, *subject*, *operator*, etc. All these words imply a natural, self-conscious action with the capacity to perform an intended action. Nothing of the kind can be attributed to artificial beings. However, the word *agent* can also be used for things, but given the possible misunderstanding that arises from calling an AI system an agent, it might make sense to call it an *artificial agent*.

tem imitates the capabilities of living beings that we can use only when we change modality from having an inherent end to *being-for-something*. An AI system does so by having a certain degree of unpredictable outputting by which it repositions itself and its actions in the framework in which it interacts with the lifeworld, as if it were a living entity. Slavery, cannibalism, animal service, and animal industry are some of the extreme variants of denying inherent ends to living beings, and this will always result in exploiting the organism's capacity for forness (the possibility to do this comes from the mutual ontological characteristic of living and non-living things that they exist as things). With technological solutions in the form of artificial agents (AI programmes, robots, unmanned vehicles, etc.), the specific capabilities are utilizable without ever risking to treat beings with inherent ends wrongly; however, AI systems as artificial agents instead of mere tools (weapons) expand their ethical relevance by having the particular "freedom" to affect the constitution of the lifeworld and because imitation alters how the life-like object affects human beings.[55] Michael C. Horowitz warns, and he is right to do so, that "AI seems much more akin to the internal combustion engine or electricity than a weapon. It is an enabler, a general-purpose technology with a multitude of applications,"[56] but it is precisely the level at which it can be utilized as a weapon with the capacity for imitating the agency of living beings that makes it suddenly important to constrain it. Why and how it is being constrained, however, is what defines the limits to the ethical and thus provokes a question concerning the ideological dimension of what has been declared "ethical," thus seemingly universal.

The following section finalizes and exemplifies the argument that there are realistic limits to having a truly ethical AI, and that these limits are fundamentally of ideological nature that may not be trumped by the current collective efforts.

[55] Humans can develop non-fictional emotions towards things, and our tendency to personify is heightened in encounters with artificial agents, especially given the tendency of developers to anthropomorphize or biomorphize their form or behaviour. See J. Blatter, E. Weber-Guskar, *Fictional Emotions and Emotional Reactions to Social Robots as Depictions of Social Agents*, "Behavioral and Brain Sciences" 2023, Vol. 46, e24, https://doi.org/10.1017/S0140525X22001716; M. Scheutz, *The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots*, in: *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. P. Lin, K. Abney, G.A. Bekey, The MIT Press, Cambridge, MA, 2012, pp. 211–214.

[56] M.C. Horowitz, *Artificial Intelligence*, op. cit., p. 39.

3. Ideological Limits to Ethical Artificial Intelligence

Regardless of the agency level of an AI system, *forness* is the attribute through which the realization of ethical AI outlines its limits. For the phenomenon of forness to manifest, an intervention in existing matter by a creator or repurpuser is required to physically or symbolically construct the object and "attach" an intention to it. It is an act, and as an act it is historically contextual: it has a cause, a reason and a purpose associated with its action in an existing cultural environment, and thus creates a direction that defines what the subjected thing will be used for in the lifeworld it will affect. Like any other technological invention, "AI development does not take place in a vacuum. The development and adoption of technology is always highly social and cultural, embedded within a rich network of human and non-human actors,"[57] and so forness is what can be monitored to reveal cultural forces that push technological solutions into motion. It is impossible not to have these elements playing a role because technical solutions do not happen outside of cultural networks. Any such network is a structural coupling of communicational systems[58] that achieve the overarching identity. Its internal consistency reveals ideology, systematized ideas that, if followed in a prescribed manner, will lead to a preferred social outcome. Through its applied forness, a technological solution can, therefore, mediate the system's congruency of behaviour to reaffirm or advance the particular social habitus.

Here, a discursive difference has to be established between ideology as present in the ethical set-up of AI, the "ethical AI", and the *ideology of AI*. The ideology of AI presupposes a systematized idea that AI systems will make the world a better place, will solve all our problems, will correct all our mistakes, will make us work less, will fulfil all our desires even before we feel them, etc.[59] This advertisement strategy, endorsed by the leading national, supranational and corporate entities, essentially depicts an image of human beings as irreparably erroneous entities that cannot be trusted and should be supplemented or replaced wherever possible to increase work efficiency. Bruce J. Berman put it aptly already in the 1990s, at

[57] L. Munn, *The Uselessness of AI Ethics*, op. cit., p. 870.

[58] A.T. Polcumpally, *Artificial Intelligence and Global Power Structure: Understanding through Luhmann's Systems Theory*, "AI & Society" 2022, Vol. 37, No. 4, pp. 1492–1493, https://doi.org/10.1007/s00146-021-012198.

[59] For an overview of ideological narratives, see L. Sias, *The Ideology of AI*, "Philosophy Today" 2021, Vol. 63, No. 3, pp. 505–522, https://doi.org/10.5840/philtoday2021514405.

a time when AI systems were a matter of science-fiction stories to the majority of the human population:

> The tendency of the AI information processing model of mind to denigrate human intellectual abilities results in what Roszak terms a "technological idolatry" that reifies the computer metaphor, generating "a haunting sense of human inadequacy and existential failure" and propagating a deference to computers "which human beings have never assumed with respect to any other technology of the past" (Roszak, 1986: 44–45). This reveals the ideological importance of AI in both legitimating and restructuring of capitalist society and generating a technological imperative requiring the installation and subordination of human labour to "intelligent" computers.[60]

Berman cited a number of influential sources showing how the possible business advantages of AI systems are related to the capitalist worldview. His findings are in agreement with a recent examination by Mikko Vesa and Janne Tienari, who demonstrate AI's elementary appeal to the "elites":

> Imagine the promise of intelligent agent programs: they never miss a detail, they never forget, and they are constantly vigilant. Nor do they (supposedly) engage in petty games nor discriminate. They appear superior in their rationality and efficiency. They do not have "agency" in any classical sense and, as a consequence, no agent-principal problems. These programs do what they are told. Only they do so a bit better every time and they transcend human capabilities in processing information many times over. Promises of superior performance or competitive advantage derived from such technologies tend to be an easy sell for decision-makers. As such, intelligent agent programs and algorithms become objects of desire in complex ways for the power elite in society. The way AI delivers competitive advantages allows for a reconfiguration of power relations. Beneath it all lies the radical promise of organizing and organizations free of human concerns and shortcomings. In effect, this creates the premise to view intelligent agent programs as perfect rational agents. However, this is largely an experiential state associated with the mastery of such code by those who control them. This promise of rationality easily positions any critique as romantic, old-fashioned, and irrational.[61]

[60] B.J. Berman, *Artificial Intelligence and the Ideology of Capitalist Reconstruction*, "AI & Society" 1992, Vol. 6, p. 111, https://doi.org/10.1007/BF02472776.

[61] M. Vesa, J. Tienari, *Artificial Intelligence and Rationalized Unaccountability: Ideology of the Elites?*, "Organization" 2022, Vol. 29, No. 6, p. 1136, https://doi.org/10.1177/1350508420963872.

Vesa and Tientari propose that "artificial intelligence functions as an ideology as it manufactures normative idea(l)s of social reality and turns these into self-evident features of discourse (Fairclough, 1989) through which we are (not) able to make sense of the world,"[62] and they attempt to explain how the approach to AI contributes to the problem of proper accountability in contemporary technology-saturated global society. The process of pushing the global civilization into an "ideological state in which power and control are exerted algorithmically" can be understood as a natural continuity of 20th-century processes initiated and organized by then-growing technocrats.[63] To give an example that helps us see beyond the danger of falling into conspiracy theories, a charter written and published by OpenAI states the following:

> OpenAI's mission is to ensure that artificial general intelligence (AGI) – by which we mean highly autonomous systems that outperform humans at most economically valuable work – benefits all of humanity.[64]

The phrase "highly autonomous systems that outperform humans at most economically valuable work" inherently implies AI's purpose, which, in turn, suggests the systematic restructuring of civilization in the context of wealth distribution. Recall that Microsoft invested $13 billion in the project behind this statement. Given the current influence of OpenAI, their mission statement confirms the sense of ideology that has been growing since the inception of technocrats. However, given the dangers of a biased and superficial approach to any examination of the clash of classes, this requires a separate analysis, and thus the outlined grand narrative is not explored further in this research. The focus is on how ideology finds its way within the ethical set-up of AI systems.

That being said, the Artificial Intelligence Act can be seen as a paradigmatic example of the systematized proscription of ideas anchored into a single phenomenon around which the phenomenon itself wants to confirm its culture. The adopted text (amend. 15, p. 9) clearly states that "development and use of ethically embedded artificial intelligence" will have to "respect Union values and the Charter." It is here that the basic ideological limit begins to show contours, because what is "ethical" is equated with "Union values." This kind of formulation demonstrates the approach to ethics as being a *preferred* set of norms, altogether

[62] Ibid., p. 1140.

[63] These processes were explained well by Maurice Duverger in 1972. See M. Duverger, *Janus. Les deux faces de l'Occident*, Fayard, Paris 1972, esp. pp. 135–247.

[64] OpenAI, *OpenAI Charter*, 9.04.2018, URL: https://openai.com/charter.

rendering the "ethical" arbitrary. It defeats the idea of a *universal ethos* practically – regardless of how much EU may claim that its values have a universal reach – and transforms the original concept of ethical as universal for every human being into a technical term. It also denies *value pluralism* as the foundation for a constructive integration of conflicting cultures, given that the same "ethical AI" in China, United States, Russia, India, Saudi Arabia, Alphabet Inc., Microsoft or the OECD will have different elements bound to the central concept. Without finding a way to overcome all sets of norms with a universal proposal, "ethical AI" may only be culturally interiorized and always completely prone to change, while the international scene of AI systems interaction will provoke cultural conflict and encourage ethics washing.[65]

Two sources can help us understand that it is not about ethical norms but about political and economic survival: national strategies and the EU social restructuring plan. Not a single national strategy of the relevant powers outside the EU, such as the United States and China, emphasizes anything other than benefits for their national gain, which from the perspective of ethics can certainly be understood as a form of ethical egoism, but in the end the harm to others is expected for the benefit of the self-oriented entity. The EU, on the other hand, is already perceived by both the EU and other political forces as an entity losing influence in the world and taking a beating in the Fourth Industrial Revolution, feeling threatened by China in particular.[66] In a special report of the Joint Research Centre on the "European perspective" on AI, it is emphasized that AI can "stimulate productivity and prosperity and lead to active work until a later age,"[67] that data is the "lifeline of Europe," and that "opening access to data and building interactions among participants is key to succeeding,"[68] presumably in the successful implementation of AI across the supranational entity for the stability of influence. In the light of this commentary, it is important to highlight an EU

[65] Cf. I. Gabriel, *Artificial Intelligence*, op. cit., p. 426.

[66] Cf. HAI, Artificial Intelligence Index Report 2023, Stanford University – Human-Centered Artificial Intelligence, Stanford 2023, URL: https://aiindex.stanford.edu/report/; B. Fricke Artificial Intelligence, 5G and the Future Balance of Power, "Konrad-Adenauer-Stiftung" 2020, No. 379, p. 6; A.T. Polcumpally, Artificial Intelligence, op. cit., p. 1498; Joint Research Centre, China: Challenges and Prospects from and Industrial and Innovation Powerhouse, Publications Office of the European Union, Luxemburg 2019, especially pp. 10–11, 20, 22, 31, 43–45.

[67] Joint Research Centre, *Artificial Intelligence: A European Perspective*, Publications Office of the European Union, Luxemburg 2018, p. 56.

[68] Ibid., p. 103.

report on the future of work, which states that "the acquisition of knowledge only through formal education will not be enough to thrive in the constantly changing world, which calls for the implementation of a lifelong-learning approach," requires "the constant re- and upskilling of workers,"[69] and a focus on "nurturing non-cognitive skills" because it "is becoming increasingly important for individuals' success in the labour market."[70]

The concept of "ethical AI" can mask the real normative for which the foundation is being developed. In the case of the Artificial Intelligence Act, the aim is gaining advantage on the global "playing field" but there is also need for risk-mitigation mechanisms for its population and reputation, and ways to overcome European national differences, as "the application of AI is often hampered by very restricted privacy laws, which make big data difficult to access."[71] Thus, it seems that the EU's behaviour confirms Hannah Arendt's claim that the social realm "is the form in which the fact of mutual dependence for the sake of life and nothing else assumes public significance."[72] For the EU's survival plan on AI to make sense, it needs to develop a fully accessible, free-flowing network of data collection equal to the networks of the United States, China, India, Russia, Japan, Australia, and other competing singularized entities, which entails not only heightened intrusion and exchange of population data but also control of the future production of data, as envisioned by the reports. The fundamental problem is that the recent progress of AI systems is due to data collected by "privacy-invasive social media applications, smartphone apps, as well as Internet of Things devices with its countless sensors."[73] Enforced regulations that supposedly regulate such data collection processes basically make no difference in practice, except to the creator of an ideological framework. These are the long-standing ethical problems of the post-privacy society, including the social and environmental costs of systematic reform, which are equally ignored and obscured by the concept of trustworthy AI.[74]

[69] Joint Research Centre, *The Changing Nature of Work and Skills in the Digital Age*, Publications Office of the European Union, Luxemburg 2019, p. 28.

[70] Ibid., p. 40.

[71] B. Fricke, *Artificial Intelligence*, op. cit., p. 5.

[72] H. Arendt, *The Human Condition*, The Chicago University Press, Chicago 1998, p. 46.

[73] T. Hagendorff, *The Ethics of AI Ethics*, op. cit., p. 110.

[74] Cf. ibid., pp. 105, 110.

In the service of the ideological system, any "ethical AI" is further diminished by the global military rivalry in which it is already assumed that "AI will give those who are well-prepared an upper hand" because "the data will enable one to 'know one's enemy as well as one knows oneself' and gain the competitive advantage."[75] The situation is so obvious that international relations and warfare experts openly discuss viable possibilities:

> Wealthy, advanced economies that have high levels of capital but also have light labor costs or small populations – middle powers such as Australia, Canada, and many European countries – often face challenges in military recruiting. For these countries, technologies that allow them to substitute capital for labor are highly attractive. […] countries can take advantage of the intersection of AI and robotics to overcome the problems caused by a small population.[76]

This creates another layer of invisible ethical problems piling up behind the idea of "ethical AI," in the sense that the broader framework of warfare remains ethically acceptable and only within this framework will questions about ethical behaviour arise. As Elke Schwarz observes, the "underlying question shifts from whether it is ethical to kill, to whether machines would do the killing better than humans. […] the ethical task at hand is to kill better and more humanely."[77] Data collection falls into the same category, as AI systems will be used to exploit the gathered information against the human source. In addition, Hagendorff emphasized that "one risk of this rhetoric is that 'impediments' in the form of ethical considerations will be eliminated completely from research, development and implementation. AI research is not framed as a cooperative global project [regardless of the emphasis in strategies on global cooperation], but as a fierce competition."[78]

Moreover, due to the limited impact of arbitrary ethical standards, those systems that insist on thorough and strict adherence to complex rules, such as the EU, may lose out in the race to win because of the rules they establish, raising the question of the moral defensibility of setting up AI systems in rigorously ethical ways. From the perspective of the population shaped by the Fourth Industrial Revolution, in the logic of racing the "ethical AI" might appear "unethical." This

[75] B. Fricke, *Artificial Intelligence*, op. cit., p. 5.

[76] M.C. Horowitz, *Artificial Intelligence*, op. cit., p. 46.

[77] E. Schwarz, *Death Machines: The Ethics of Violent Technologies*, Manchester University Press, Manchester 2019, p. 165.

[78] T. Hagendorff, *The Ethics of AI Ethics*, op. cit., p. 107.

question is underpinned by another concept with which political and corporate entities try to profit from the development of AI and limit its ethics: "hampering" of progress. For example:

> the right to explanation in the GDPR will come at cost in the efficiency or efficacy of the AI systems in question: optimisation and efficiency will be partially sacrificed for increases in transparency and accountability. While this is unproblematic in itself, as critics of regulation like to point out, such initiatives decrease the competitiveness of such systems on the global market, thus diminishing their likely overall representation and impact at the global level.[79]

This issue is linked to another problem on the level of the ethically aligned design of AI – the fact that the ethical properties which we would like AIs to have, such as transparency and explainability, may, on the one hand, prevent the development of highly efficient AI systems that find correlations "in data too huge for human to assess,"[80] and, on the other, lessen the possibility of non-human "intelligent" behaviour leading to new discoveries.

The general framework of ideological limits applies to all the listed categories of ethical AI, but already at the level of design and implementation experts are familiar with the so-called *inclusive design paradox*, where "positively improving a system to include as many values as possible might negatively influence the overall application,"[81] creating too many competing principles for the AI system to resolve it appropriately for everyone. Joris Krijger called this the effect of *inter-principle tension*, "the challenge of implementing multiple values and principles in one design," to which he added *intra-principle tension*, "the challenge of translating a single normative principle (in)to a specific technological design."[82] His division can be updated with the notion of *extra-principle tension*, which can be understood as the challenge of resolving competing norms between what is included in the AI system and what has been excluded. An AI system which by necessity has to adhere to particular values will enforce the subjugation to these values in every situation in which it finds itself. Little is known about what hap-

[79] H.-Y. Liu, *The Power Structure of Artificial Intelligence*, "Law, Innovation and Technology" 2018, Vol. 10, No. 2, p. 206, https://doi.org/10.1080/17579961.2018.1527480.

[80] H.M. Roff, *Artificial Intelligence*, op. cit., p. 137.

[81] J. Krijger, *Enter the Metrics: Critical Theory and Organizational Operationalization of AI Ethics*, "AI & Society" 2022, Vol. 37, No. 4, p. 1432, https://doi.org/10.1007/s00146-021-01256-3.

[82] Ibid.

pens when ethical AI encounters different types of norms, and yet no thorough research has been conducted by policy makers working on strategies and regulation, while scholars have only began to explore in more depth how communication between AI systems in a saturated environment should be processed.

Of the major problems related to the ideological elements of "ethical AI," last but not least is the demographic structure of those involved in the development and discussion on AI, dominated by the white male population with some common characteristics related to the underlying cultural and, possibly, biological traits. The tech-culture toxicity goes beyond science and business solutions,[83] as AI systems are heavily present in the video-gaming industry in which the past ten years were abundant with sexual, racial, and exploitation scandals, as well as labour abuse, usually in the working environment of leading giants such as Activision Blizzard, CD Project Red, and Electronic Arts. It is a "culture known for the hypermasculine coder or 'brogrammer,'" where "60% of women reported unwanted sexual advances."[84] Recently, a class action lawsuit that "has accused a widely celebrated tech company of fostering racist conditions for years, including daily subjection to racial slurs, being assigned menial jobs in a segregated area of the factory, and being passed over in promotions for management."[85] This is the same social circle that systematically ignores the application of ethical principles, as pointed out in section 1. Male-normative values are most evident in the domain of ethical design and ethical development of AI, particularly in the male approach to understanding AI and solving problems. Classical empirical studies show that "women do not, as men typically do, address moral problems primarily through a 'calculating,' 'rational,' 'logic-oriented' ethics of justice, but rather interpret them within a wider framework of an 'empathic', 'emotion-oriented' ethics of care."[86] However:

> In AI ethics, technical artefacts are primarily seen as isolated entities that can be optimised by experts so as to find technical solutions for technical problems. What is often lacking is a consideration of the wider context and the comprehensive relationship networks in which technical systems are embed-

[83] For an abundance of examples and the history of this approach, see S. Watcher-Boettcher, *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*, W.W. Norton & Company, New York 2017.

[84] L. Munn, *The Uselessness of AI Ethics*, op. cit., p. 871.

[85] Ibid.

[86] T. Hagendorff, *The Ethics of AI Ethics*, op. cit., p. 103.

> ded. In accordance with that, it turns out that precisely the reports of AI Now (Crawford et al. 2016, 2019; Whittaker et al. 2018; Campolo et al. 2017), an organization primarily led by women, do not conceive AI applications in isolation, but within a larger network of social and ecological dependencies and relationships (Crawford and Joler 2018), corresponding most closely with the ideas and tenets of an ethics of care (Held 2013).[87]

In order to understand the extent to which the systems actually contribute to improving the quality of life of the population, we need to pay attention to who exactly is developing the framework for the use of AI, what characteristics and problems a particular user group has, what kind of approach they have to their discoveries or inventions, and whether they published anything that may clearly explain their motives. For example, in the context of social instability in the United States in relation to policing and minority populations, AI prediction or identity recognition systems simply cannot be deployed as isolated technical support akin to patrol vehicles, security cameras or emergency call networks, unless there is a specific agenda to embolden the ongoing stratification, because they are based on data reflecting past social practices riddled with racial behaviour and corruption.

If we look at AI systems as conveying technological nodes within a social system, then we can identify the kinds of cultural contexts that are attached to them and the proscriptions they mediate, that is, the system of ideas they indirectly represent or endorse. We then come to understand how they can behave as springboards for aims beyond their internal ethical set-up. Following the differentiation given at the beginning of sections 2 and 3, it can be concluded that ideological elements that affect ethical set-ups in AI systems and thus, by generating ideological bias in realistic deployment, limit what the "ethical" can achieve, manifest at three levels:

– social framework, endorsing AI systems development and application, which can be studied to identify the broadest forces and the most important actors within each social subnetwork endorsing AI systems to determine why they are being forced upon the citizens and what general claims and arguments for these claims are attached to the agenda;
– engineering framework, which can be studied to identify what cultural values were endorsed or implemented under the presented set of principles, as is the case with the EU, which wants AI systems to embody "Union values"

[87] Ibid., pp. 103–104.

specifically, or OpenAI, which exploited legal loopholes and economic instability;

- use of AI systems, which does not constrain ethical AI internally but externally, and can thus be studied in comparison to what the AI system was designed for, to grasp the exploitation, corruption or deviation of its ethical set-up, which is oblivious to social context, such as the deployment of AI systems in warfare, policing, and legal disputes.

Any formulation of ethical guidelines implies communication with ideological frameworks. However, many participants in the field of AI overlook the presence of ideological elements during the development, deployment and use of AI systems, and the fact that ethical AI is ethical only insofar as what stands for "ethical" is either universally acceptable or does not attempt to push an agenda. The situation with AI development is quite the opposite – it has become entangled with the ideological framework stemming from political and economic interests, and the practices surrounding the concept of ethical AI already show that the concept can serve as a tool of manipulation. The infusion of ideological elements into ethical regulations, which will eventually align with legal systems and gain social acceptance, needs to be examined in more depth.

4. Conclusion

There may be ways to perfect the design, development and behaviour of AI systems that support humanity in its evolution of the humane and resemble acceptable moral behaviour or an appropriate universal code of conduct, but the deployment of AI systems is the dimension where their utility is encumbered by the broader ideological framework that arises from the cultural conflicts and habits that have historically been in place. The exploration, discussion and development of "true" ethical AI is what would inevitably "hinder" the developmental progress of AI systems, as it would "impede" the particularist and exceptionalist political and economic agenda, which is certainly one of the reasons why the issue is systematically avoided or overlooked. But the problems I have highlighted and discussed in this paper, undoubtedly not the only ones plaguing AI, will persist alongside everything that will unfold with AI systems in the near future. In terms of how AI could genuinely contribute to humanity, "business and politics as usual" is the worst realization of its potential because it fails to address the stalled course of civilizational development encumbered by conflicts, low quality of life, and resource depletion.

The main issue of justifying the fundamental idea that the use of structurally saturated AI systems is imperative for the "better future of humanity" is argumentatively buried by the positivist and pragmatic approach to AI, and that too is ideological in nature: there is a lacuna between using AI systems to increase the efficiency of personal endeavours or prevent harm, and building a global infrastructure to monitor the conversion of each of our atoms into fuel for the survival of the current framework. In order for us to understand what the "ethical" in ethical AI presupposed, it has to be deconstructed to its fundamental components. For example, the EU's Artificial Intelligence Act continuously emphasizes that AI systems have to have an ethical set-up and have to be regulated by law but these constraints should not hamper their development. This means that in any arbitrarily evaluated moral dilemma, the developmental breakthrough always prevails over the moral constraints. Thus, for example, if achieving the EU's goals in the AI race necessitates gathering extensive information about citizens and a systematic restructuring of their lives, we can expect that the EU will bypass inconvenient regulations, such as privacy laws, and trample over the unregulated. However, it will be doing so to perpetuate the existing political and economic system – a system that Europeans themselves shaped over the past 200 years – and not to change the course of European citizens' existence and foster a better approach to life. Because AI systems appeared in an epoch of major ideological conflicts, as technological inventions they must be interpreted as the possible mediators of ideological goals, meaning that the content of the ethical in "ethical AI" has its boundaries drawn by ideological elements defining the product.

The considerations presented in this paper were limited to drawing attention to the ways in which ideological elements can enter the ethical set-up, which, based on its name alone, is often misguidedly represented or thought of as universal. I aimed to show that even a quite transparent, straightforward approach to presenting ethical AI, such as that of the EU in the Artificial Intelligence Act, presupposes aims and limitations to its applicability that subdue the ethical principles selected to form the ethical set-up of AI. These aims and limitations belong to the broader ideological frameworks that become attached to AI systems. Further steps that can be taken to broaden the research are a closer inspection of how ideological elements come into play at each designated level, followed by sequential case studies, and an attempt to develop and demonstrate a toolkit for identifying ideological bias.

Bibliography

Allen C., Wallach W., *Moral Machines: Contradiction in Terms or Abdication of Human Responsibility?*, in: *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. P. Lin, K. Abney, G.A. Bekey, The MIT Press, Cambridge, MA, 2012, pp. 55–67.

Alonso E., *Actions and Agents*, in: *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish, W.M. Ramsey, Cambridge University Press, Cambridge 2014, pp. 232–246.

Arendt H., *The Human Condition*, The Chicago University Press, Chicago 1998.

Baez B., *Technologies of Government: Politics and Power in the "Information Age"*, Information Age Publishing, Charlotte 2014.

Beck U., *Gegengifte. Die organisierte Unverantwortlichkeit*, Suhrkamp Verlag, Frankfurt am Main 1988.

Berman B.J., *Artificial Intelligence and the Ideology of Capitalist Reconstruction*, "AI & Society" 1992, Vol. 6, pp. 103–114, https://doi.org/10.1007/BF02472776.

Blatter J., Weber-Guskar E., *Fictional Emotions and Emotional Reactions to Social Robots as Depictions of Social Agents*, "Behavioral and Brain Sciences" 2023, Vol. 46, e24, https://doi.org/10.1017/S0140525X22001716.

Boddington P., *AI Ethics: A Textbook*, Springer, Singapore 2023.

Campolo C., Crawford K., *Enchanted Determinism: Power without Responsibility in Artificial Intelligence*, "Engaging Science, Technology, and Society" 2020, Vol. 6, pp. 1–19.

Carter W.A., Crumpler W.D., *Smart Money on Chinese Advances in AI: A Report of the CSIS Technology Policy Program*, Center for Strategic and International Studies, Washington 2019.

Chitty N., Dias S., *Artificial Intelligence, Soft Power and Social Transformation*, "Journal of Content, Community and Communication" 2017, Vol. 6, No. 3, pp. 1–14.

Council of the European Union, *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: Analysis of the Final Compromise Text with a View to Agreement*, no. Cion doc. 8115/21, Brussels, 26 January 2024.

De Gregorio G., *The Normative Power of Artificial Intelligence*, "Indiana Journal of Global Legal Studies" 2023, Vol. 55, pp. 1–19, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4436287.

DiMatteo L.A., *Artificial Intelligence: The Promise of Disruption*, in: *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, eds. L.A. DiMatteo, C. Poncibò, M. Cannarsa, Cambridge University Press, Cambridge 2022, pp. 3–17.

Duverger M., *Janus. Les deux faces de l'Occident*, Fayard, Paris 1972.

Ekmekci P.E., Arda B., *Artificial Intelligence and Bioethics*, Springer Nature Switzerland, Cham 2020.

European Parliament, P9_TA(2023)0236: Artificial Intelligence Act, June 2023.

Fair Trials, *Automating Injustice: The Use of Artificial Intelligence and Automated Decision-Making Systems in Criminal Justice in Europe*, 9.09.2021, URL: https://www.fairtrials.org/articles/publications/automating-injustice/.

Franklin S., *History, Motivations, and Core Themes*, in: *The Cambridge Handbook of Artificial Intelligence*, eds. K. Frankish, W.M. Ramsey, Cambridge University Press, Cambridge 2014, pp. 15–33.

Fricke B., *Artificial Intelligence, 5G and the Future Balance of Power*, "Konrad-Adenauer-Stiftung" 2020, Vol. 379, pp. 1–9.

Gabriel I., *Artificial Intelligence, Values, and Alignment*, "Minds and Machines" 2020, Vol. 30, No. 3, pp. 411–437, https://doi.org/10.1007/s11023-020-09539-2.

Hagendorff T., *The Ethics of AI Ethics: An Evaluation of Guidelines*, "Minds and Machines" 2020, Vol. 30, No. 1, pp. 99–120, https://doi.org/10.1007/s11023-020-09517-8.

HAI, *Artificial Intelligence Index Report 2023*, Stanford University – Human-Centered Artificial Intelligence, Stanford 2023, URL: https://aiindex.stanford.edu/report/.

High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, European Commission, European Union 2019.

Horowitz M.C., *Artificial Intelligence, International Competition, and the Balance of Power*, "Texas National Security Review" 2018, Vol. 1, No. 3, pp. 36–57.

IEEE Global Initiative, *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*, 2019, URL: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf.

Introna L.D., Wood D., *Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems*, "Surveillance and Society" 2004, Vol. 2, No. 2–3, pp. 177–198.

Joint Research Centre, *Artificial Intelligence: A European Perspective*, Publications Office of the European Union, Luxemburg 2018.

Joint Research Centre, *China: Challenges and Prospects from an Industrial and Innovation Powerhouse*, Publications Office of the European Union, Luxemburg 2019.

Joint Research Centre, *The Changing Nature of Work and Skills in the Digital Age*, Publications Office of the European Union, Luxemburg 2019.

Krijger J., *Enter the Metrics: Critical Theory and Organizational Operationalization of AI Ethics*, "AI & Society" 2022, Vol. 37, No. 4, pp. 1427–1437, https://doi.org/10.1007/s00146-021-01256-3.

Kroes P., *Technical Artefacts: Creations of Mind and Matter*, Springer, Dordrecht 2012.

Li P., Lang J., Islam M.A., Ren S., *Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models*, arXiv:2304.03271 [cs.LG], https://doi.org/10.48550/arXiv.2304.03271.

Liao S.M., *A Short Introduction to the Ethics of Artificial Intelligence*, in: *Ethics of Artificial Intelligence*, ed. S.M. Liao, Oxford University Press, Oxford 2020, pp. 1–42.

Liboreiro J., *European Regulators Put Microsoft's $13 Billion Bet on OpenAI under Closer Scrutiny*, EuroNews, 9.01.2024, URL: https://www.euronews.com/my-europe/2024/01/09/european-regulators-put-microsofts-13-billion-bet-on-openai-under-closer-scrutiny.

Liu H.-Y., *The Power Structure of Artificial Intelligence*, "Law, Innovation and Technology" 2018, Vol. 19, No. 2, pp. 197–229, https://doi.org/10.1080/17579961.2018.1527480.

Loke S.W., *Designed to Cooperate: A Kant-Inspired Ethic of Machine-to-Machine Cooperation*, "AI and Ethics" 2022, Vol. 3, No. 3, pp. 991–996, https://doi.org/10.1007/s43681-022-00238-5.

Magnani L., *Morality in a Technological World: Knowledge as Duty*, Cambridge University Press, Cambridge 2007.

Marcellis-Warin N. de, Marty F., Thelisson E., Warin T., *Artificial Intelligence and Consumer Manipulations: From Consumer's Counter Algorithms to Firm's Self-Regulation Tools*, "AI and Ethics" 2022, Vol. 2, No. 2, pp. 259–268, https://doi.org/10.1007/s43681-022-00149-5.

Matz S.C., Kosinski M., Nave G., Stillwell D.J, *Psychological Targeting as an Effective Approach to Digital Mass Persuasion*, "PNAS" 2017, Vol. 114, No. 48, pp. 12714–12719, https://doi.org/10.1073/pnas.1710966114.

McQuate S., *UW Researcher Discusses Just How Much Energy ChatGPT Uses*, University of Washington, 27.07.2023, URL: https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/.

Melkevik Å., *The Internal Morality of Markets and Artificial Intelligence*, "AI and Ethics" 2023, Vol. 3, No. 1, pp. 113–122, https://doi.org/10.1007/s43681-022-00151-x.

Munn L., *The Uselessness of AI Ethics*, "AI and Ethics" 2023, Vol. 3, No. 3, pp. 869–877, https://doi.org/10.1007/s43681-022-00209-w.

Nduka C., *How OpenAI Transitioned from a Nonprofit to a $29B For-Profit Company*, Hackernoon, 28.03.2023, URL: https://hackernoon.com/how-openai-transitioned-from-a-nonprofit-to-a-$29b-for-profit-company.

Novo C., *The Water Cost of Artificial Intelligence Technology*, "SmartWaterMagazine," 12.09.2023, URL: https://smartwatermagazine.com/news/smart-water-magazine/water-cost-artificial-intelligence-technology.

Nyholm S., *The Ethics of Crashes with Self-Driving Cars: A Roadmap, I*, "Philosophy Compass" 2018, Vol. 13, No. 7, e12507, pp. 1–10, https://doi.org/10.1111/phc3.12507.

OpenAI, *OpenAI Charter*, 9.04.2018, URL: https://openai.com/charter.

Parkinson J., Bariyo N., Chin J., *Huawei Technicians Helped African Governments Spy on Political Opponents*, "Wall Street Journal," 15.08.2019, URL: https://www.wsj.com/articles/huawei-technicians-helped-african-governments-spy-on-political-opponents-11565793017#comments_sector.

Perrigo B., *OpenAI Used Kenyan Workers on Less Than $2 per Hour to Make ChatGPT Less Toxic*, "Time," 18.01.2023, URL: https://time.com/6247678/openai-chatgpt-kenya-workers/.

Polcumpally A.T., *Artificial Intelligence and Global Power Structure: Understanding through Luhmann's Systems Theory*, "AI & Society" 2022, Vol. 37, No. 4, pp. 1487–1503, https://doi.org/10.1007/s00146-021-012198.

Roff H.M., *Artificial Intelligence: Power to the People*, "Ethics and International Affairs" 2019, Vol. 33, No. 2, pp. 127–140, https://doi.org/10.1017/S0892679419000121.

Scheutz M., *The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots*, in: *Robot Ethics: The Ethical and Social Implications of Robotics*, eds. P. Lin, K. Abney, G.A. Bekey, The MIT Press, Cambridge, MA, 2012, pp. 205–221.

Schwarz E., *Death Machines: The Ethics of Violent Technologies*, Manchester University Press, Manchester 2019.

Seaton H., *The Construction Technology Handbook*, John Wiley & Sons, Hoboken 2021.

Sias L., *The Ideology of AI*, "Philosophy Today" 2021, Vol. 63, No. 3, pp. 505–522, https://doi.org/10.5840/philtoday2021514405.

Simonite T., *What Really Happened When Google Ousted Timnit Gebru*, Wired, 8.06.2021, URL: https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/.

UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, Paris 2022.

Verbeek P.-P., *Moralizing Technology: Understanding and Designing the Morality of Things*, The University of Chicago Press, Chicago 2011.

Vesa M., Tienari J., *Artificial Intelligence and Rationalized Unaccountability: Ideology of the Elites?*, "Organization" 2022, Vol. 29, No. 6, pp. 1133–1145, https://doi.org/10.1177/1350508420963872.

Vries A. de, *The Growing Energy Footprint of Artificial Intelligence*, "Joule" 2023, Vol. 7, No. 10, pp. 2191–2194, https://doi.org/10.1016/j.joule.2023.09.004.

Wagner B., *Ethics as an Escape from Regulation: From "Ethics Washing" to Ethics-Shopping?*, in: *Being Profiled: Cogitas Ergo Sum. 10 Years of Profiling the European Citizen*, eds. I.E. Bayamlıoğlu, I. Baraliuc, L. Janssens, M. Hildebrandt, Amsterdam University Press, Amsterdam 2018, pp. 84–89.

Watcher-Boettcher S., *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*, W.W. Norton & Company, New York 2017.

Xiang C., *"He Would Still Be Here": Man Dies by Suicide after Talking with AI Chatbot, Widow Says*, Vice, 30.03.2023, URL: https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says.

Yampolskiy R.V., *Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach*, in: *Philosophy and Theory of Artificial Intelligence*, ed. V.C. Müller, Springer-Verlag, Berlin 2013, pp. 389–396.