The Main Challenges of AI Ethics: Historical Contextualization, Black-Boxing, Social Biases, Labor Invisibility

Konstantinos Konstantis

Department of History and Philosophy of Science National and Kapodistrian University of Athens, Athens, Greece
konstkon@phs.uoa.gr

Abstract

In the research described here, I argue that an adequate approach to AI ethics should include the four topics below. My aim is to answer the question of which are the necessary topics that someone should have under consideration in order to make an adequate approach to AI ethics. First, a critical history of AI, which focuses not on the technical differentiations between previous and following technologies, but on the social, economic, and political context in which artificial intelligence is designed, developed, and used. Second, an overview of the issues that most of the time are described as AI ethics, such as fairness, accountability, and transparency, in order to have the ability to understand what is missing from these approaches. A study on the black box of AI is necessary, not only from a technical perspective, but mainly from a perspective that is directly related to the political, social, and economic reasons that enforce and reinforce this black box, revealing, among others, the social relations, the hidden labor, and the "unintelligence" that are hidden under this black box. Third, an analysis of specific cases through critical approaches which take into account capitalism, with all the social, political, and economic relations that are connected with it. In this way, the emergence of biases, inequalities, and discriminations, becomes not a bag, but the substance of AI. Fourth, a study on the hidden labor of AI and the concerns regarding the future of work and AI. The study on hidden labor which is related with AI, is important in order, first, to criticize the intelligence and autonomy of AI systems, and second, to make visible the terrible working conditions of some workers, as a try to change them. The discussion regarding the future of work should not only contain discourses regarding the circular function of capitalism or vague ideas about ethical implementations of AI in the workplace. An adequate discussion should take into account the social, political, and economic relations of our society and ultimately challenge the current form of capitalism. I argue that all the above should be included in an adequate study of AI ethics.

Historical Contextualization

AI should not be perceived only as technological artifact, and therefore, an adequate study of AI should include perspectives from many scientific fields. Artificial intelligence can be seen as a branch of political science that is conveyed through computer science. What is needed, is to reveal the social orders that are presented as a priori reified in our society, under the cloak of AI (Penn 2020). Artificial intelligence was mechanical and electrical long before it became electronic. Computing machines are presented as intelligent, regardless the time of period, in order for the computing machines to be presented as the source of value. In this way, the manual and intellectual labor that is needed for the design, development, and use of a machine to be presented as intelligent, is concealed, and the computing wages were decreased. In this way, the accumulation of surplus computing value is feasible (Tympas 2017).

For a critique of the idea of "smartness", it is crucial to criticize the dominant ideology supporting that every aspect of people's lives could and should be under algorithmic manipulation in order for "smartness" to solve every crisis, such as environmental degradation, inequality, and injustice, and not only research the way that technology developed and used (Halpern, Mitchell, and Geoghegan 2017).

Black Boxing and Social Biases

Privacy and surveillance, accountability, manipulation of behavior, biases, automation and unemployment, opacity, and fairness are only some of the AI ethics issues that are in the spotlight (Müller 2020). There is also a variety of discussions regarding superintelligence and robot ethics (Coeckelbergh 2022).

I argue that in order to try to tackle these challenges, first, what is necessary is the opening of the black box of AI. In order to open sufficiently the black box of AI, an STS (Science and Technology Studies) approach is needed. This approach reveals, first, that AI can be opaque even to its designers (Burrell 2016), due, not only to technical factors, but as a technology, to a combination of social, economic, and political factors too, second, that AI can conceal secret algorithms in favor of big tech (Pasquale 2015), third, that AI can be presented as the solution to social problems, despite the fact that this could not be the case (Broussard 2018), fourth, that a "human-centered design" for AI could not be enough for an ethical design of it (Benjamin 2019), fifth, that AI could be examined not only as a technology, but also as discourses, a concept, a field, and a set of practices, which reinforce the social white supremacy (Katz 2020), and sixth, that under the black box of AI, is not concealed (only) the idea of imitating the human's biological intelligence, but the

intelligence of labor and social relations (Pasquinelli 2023). An overview of specific cases is necessary in order to understand that biases are not a bag, but a structural element of AI. These cases disclose, first, that the inequality that dominates in many forms of people's lives, is reinforced by AI systems (Wiggins and Jones 2023), second, that AI through the combination of algorithms and mathematics, could be considered even as a weapon that would reinforce inequalities and discrimination (O'Neil 2016), and third, that in order to avoid biases, trying for better data is not enough, because biases in AI represent the biases of society and therefore there are no neutral data (Eubanks 2018).

Labor Invisibility

Thomas S. Mullaney, one of the editors of the book Your Computer Is on Fire, argues that nothing is virtual. Everything that is presented as taking place in a virtual, online world in an automated way, includes human labor, which is hidden and kept invisible (Mullaney 2021). Ensmenger supports that perceiving "Cloud" as a factory gives the chance to understand better the notion of the place in the information economy, the labor needed for its operation, and appreciate the importance of infrastructure (Ensmenger 2021). Kate Crawford argues that artificial intelligence is neither artificial nor intelligence. She supports the idea that artificial intelligence is a "registry of power", including, among others, social practices, politics, and culture, and therefore, is not a purely technical domain. She also highlights the fact that workers are needed both for building and maintaining the infrastructures that AI is needed and for testing AI systems. Despite the fact that these workers are necessary for AI, they are underpaid. Instead of asking whether AI will replace humans (Konstantis et al. 2024), Crawford offers a different perspective. She is interested in how humans are treated like robots in the field of work, and what does this mean for themselves and their labor (Crawford 2021).

Transforming Theoretical into Practical Research (Planned Work)

In my research, I offer a combination of the perspectives mentioned above. However, this combination includes only secondary and primary literature material.

What is missing is a combination of theoretical research with practical ways on how to open the black box of AI. Working together with engineers, I will study how AI could have been designed and developed differently, if engineers had been taking into consideration the STS perspectives mentioned above.

Acknowledgments

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 3rd Call for HFRI PhD Fellowships (Fellowship Number: 5188).

References

Benjamin, R. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK, Medford, Massachusetts: Polity.

Broussard, M. 2018. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, Massachusetts, London, England: The MIT Press.

Burrell, J. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society*, 3(1): 1–12.

Coeckelbergh, M. 2022. *Robot Ethics*. Cambridge, Massachusetts, London, England: The MIT Press.

Crawford, K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven and London: Yale University Press.

Ensmenger, N. 2021. The Cloud Is a Factory. In Mullaney, T. S.; Peters, B.; Hicks, M.; and Philip, K., eds., *Your Computer Is on Fire*, 29–49. Cambridge, Massachusetts, London, England: The MIT Press.

Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Halpern, O.; Mitchell, R.; and Geoghegan, B. D. 2017. The smartness mandate: Notes toward a critique. *Grey Room*, 68: 106–129.

Katz, Y. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press.

Konstantis, K.; Georgas, A.; Faras, A.; Georgas, K.; and Tympas, A. 2024. Ethical considerations in working with ChatGPT on a questionnaire about the future of work with ChatGPT. *AI and Ethics*, 4: 1335–1344.

Mullaney, T. S. 2021. Your Computer Is on Fire. In Mullaney, T. S.; Peters, B.; Hicks, M.; and Philip, K., eds., *Your Computer Is on Fire*, 3–9. Cambridge, Massachusetts, London, England: The MIT Press.

Müller, V. C. 2020. Ethics of Artificial Intelligence and Robotics. https://plato.stanford.edu/archives/win2020/entries/ethics-ai/.

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, Massachusetts, London, England: Harvard University Press.

Pasquinelli, M. 2023. *The Eye of the Master: A Social History of Artificial Intelligence*. London, UK, Brooklyn, New York: Verso.

Penn, J. 2020. *Inventing Intelligence: On the History of Complex Information Processing and Artificial Intelligence in the United States in the Mid-Twentieth Century*. Ph.D. thesis, University of Cambridge.

Tympas, A. 2017. *Calculation and Computation in the Pre-electronic Era*. New York, NY: Springer.

Wiggins, C.; and Jones, M. L. 2023. *How Data Happened: A History from the Age of Reason to the Age of Algorithms*. New York, NY, London, UK: W. W. Norton & Company.