REVIEW ARTICLE

Mapping the Road of the Ethical Dilemmas Behind Artificial Intelligence

Georgios L. Vousinas (iD)
National Technical University of Athens, Greece.

Ilektra Simitsi (iD)
University of the Peloponnese, Greece.

Georgia Livieri (iD)
Cyprus University of Technology, Cyprus.

Georgia - Chara Gkouva (iD)
Panteion University of Social and Political Sciences, Greece.

Iris - Panagiota Efthymiou (iD)
University of Piraeus, Greece.

Abstract

This paper aims to provide a roadmap on the major ethical dilemmas behind the application of AI, hopefully initiating a fruitful discussion on this crucial matter that is at the forefront of scientific and global attention. To achieve this the main ethical dilemmas are identified by the most recent literature review, and then are briefly analysed to provide a holistic approach on the ethical issues that arise by the use of AI systems. Furthermore, the authors provide some best practices for dealing with the existing and future challenges of AI.

Keywords: Artificial Intelligence, Ethics, Ethical dilemmas, Human judgement, Data

1. Introduction

The study of Artificial Intelligence (AI) ethics is not something new in the global scientific agenda, but what falls under this subject depends on the term AI is defined. For this paper we will use the following widely accepted definition: "AI is defined as a system's ability to interpret external data correctly, to learn from such data and to use those learnings to achieve certain goals and tasks through flexible adaptation" (Kaplan and Haenlein, 2019). AI has numerous benefits and aims to help in addressing urgent global challenges thus, providing a solution for accomplishing the United Nations' Sustainable Development Goals (UN SDGs). On the other hand, AI is also related with important and difficult to calculate drawbacks for society, especially regarding its ethical side. The core features of

AI that can raise ethical concerns are the capability of learning and acting more or less in an autonomous way on the grounds of external input and adaptation. The debate around AI ethics has become revitalized in recent years driven by recent developments and achievements of various AI techniques, and their extensive application in sectors such as smart cities (Ryan and Gregory, 2019; Ryan 2019b), smart learning environments (Efthymiou-Egleton, 2020), agriculture (Ryan 2019a) and transportation (Ryan 2020). As a result, the main objective of this paper is to provide a roadmap on the major ethical dilemmas behind the implementation of AI, clarifying and analysing them, and hopefully initiating a fruitful discussion on this crucial matter that is at the forefront of scientific and global attention.

2. On the Ethical Side of AI: A Literature Review

In spite of the fact that AI systems are considered a great promise for humanity, their application can also create significant ethical challenges (Winfield, 2019). Considering their broad usage in solving complicated issues to several domains such as healthcare, the financial sector and military, there is a growing worry concerning the replacement of human workforce by technology e.g., robots., a fact that will lead to high levels of unemployment in the foreseeable future (Frey and Osborne, 2013).

Additionally, one of the main arguments against AI is that it promotes the concentration of power in the hands of a few privileged (Nemitz, 2018), while creating inequality of wealth distribution (Bryson, 2019). Likewise, the rapid evolution of AI could raise discrimination issues for example by forcing people without sufficient technical skills out of the labor market thus, widening the existing socioeconomic gap, hence affecting the disadvantaged groups such as the minorities, women etc.

Even more, as the usage of autonomous robots and unmanned vehicles is ever growing, and the former have replaced operations that prior required human involvement, the issue of manipulation of human judgment and behaviour comes to surface. For example, major ethical dilemmas are raised regarding the operation of autonomous weapons systems (AWS) as they are designed to bring about physical harm and act autonomously thus, beyond the human judgement and/or control leading to a serious violation of fundamental human rights. Likewise, autonomous vehicles (AVs) are often called to decide among two evils e.g., which life to save, the driver's or the pedestrian's? (Bonnemains et al., 2018). This autonomous operation is also related with severe mistakes, the so-called AI stupidity, the treatment of which is crucial especially when human life is at stake. Since AI technologies are developed by humans, they are also prone to bias which can stem from the data used to train them as in the case of AI systems that exploit a database of only specific demographic groups e.g., favour men

instead of women. As a result, control over AI is a key issue in the process of designing and implementing such innovating systems.

While the advancement of technology makes AI systems even more intelligent and autonomous there is always the fear that, in the near future, they will surpass human ability and take control over humans with devastating results for all mankind. In the same direction, much concern is also about the issues of security, data protection and privacy, with the main question been how to balance among the positive effects of AI i.e., accessibility and transparency, with data privacy and security issues. And when we have to do with safety-critical systems such as AWS and medical diagnosis mechanisms, security and privacy are of utmost priority.

Misinformation and fake news are also considered key issues of the AI ethics analysis as recent history has shown that AI-powered systems have been widely exploited to manipulate people such as during election procedures (Bradshaw, S. & Howard, P., 2017). A typical example is the well-known Cambridge Analytica scandal that involved the illegal collection (via Facebook) of the psychological profiles of American voters in order to affect their personality traits (Efthymiou-Egleton et al., 2020).

Finally, all the previous ethical dilemmas behind AI adoption are of major significance when we refer to kids. AI-powered systems must take into consideration the kids' unique traits e.g., in the design of AI-enabled toys and must ensure children's safety, support their development and well-being, protect their data and privacy, and provoke fairness.

3. Analysis of the Main Ethical Dilemmas Behind AI

This section aims to provide a brief analysis of the previously identified ethical dilemmas by the literature and discuss their effects in the implementation of AI.

3.1. Unemployment due to Automation

The way AI works is impressive in terms of quality, offering flexibility and quantity. Nevertheless, due to this development, many professions will be lost or at least modified, especially those characterized by routine tasks and duplication e.g., industrial assembly occupations and non-executive office jobs (Koski, O., 2018). Furthermore, if the employees training is not well implemented, social inequalities, unemployment and social unrest will increase. Finland is one of the first countries that implemented AI trainings so as to equip workers, unemployed and entrepreneurs, with education related to their positions, wages and job upgrades (Koski, O., 2018). But what about the other countries? Can they adjust to the change that AI brings, or will they replace their staff with machines?

3.2 Discrimination Issues & Inequality of Wealth Distribution

The discrimination concerns on the basis of skin color, gender or racial origin regarding the use of AI are crucial due to their effects on people. AI can be used in various policies, from unemployment benefits and interest rates to crime prevention, selection of employees and advertising (Gerards & Zuiderveen Borgesius, 2020). Hence, a manipulated AI-driven price may lead to a higher paying rate for a specific societal group. To be noted that there is intentional discrimination, apart from unintentional, where for a specific reason an organization may perform a hypothesis for the target group for their own profit (Duhigg, 2012). To address these discrimination risks there are developed institutional human rights monitoring bodies, such as the Council of Europe and the European Commission against Racism and Intolerance.

3.3 Manipulation of Human Judgment & Behaviour

In case of AI, the process of affecting the preferences of the consumers and voters is more evident than ever due to the influence and instrumentalization of network systems such as social media, which not only hinders the development of personality and emotional intelligence of individuals, but also leads them in specific directions in order to make profit (buy) or to pursue political interests (fellow-minded citizens, voting option). Is AI smarter than its creator and could judge what is right, moral and more legitimate than the emotional intelligence and judgment of individuals? And if so, to what extend is this approach democratic? Furthermore, how do we know that the algorithm works properly after collecting the data of the same people? (Silberg J. And Manyika, J, 2019). There are many control mechanisms such as AI institutes and organizations that study and observe the evolution of behaviorism and how to prevent from the emerging risks.

3.4 Roboethics

In the previous years there has been a growing interesting on the potential impact of future AI systems as well as robotics, while leading thinkers around the globe have publicly warned regarding the risk of a dystopian future, as the degree of complexity of such systems further evolves (Torresen, 2018). In the near future, it is estimated that robots and autonomous systems will have a widespread exploitation in society, including self-driving vehicles and service robots at both work and home, posing the question: "How quickly will we see a transformation?". In fact, the technological shift from industrial to service robots, represents a progress into more personalized as well as more autonomous systems (Torresen, 2018). Our society is dealing with potential challenges from future highly intelligent systems regarding jobs and technology risks. In conclusion, although new technology brings about

many gains, its misuse is always a greater threat compared to the technology itself getting totally out of control.

3.5 Treatment of AI Stupidity - Mistakes of AI Bias

As intelligence grows through learning, the algorithms and machines shall get trained in order to "learn". One approach to achieve that is through deep learning where machines learn the best patterns and how to precisely classify data or the approach of reinforcement learning where machines are "rewarded" if they achieve a goal (Siegfried, 2020). After feeding the system with the adequate input, it is tested towards its performance so to undertake the necessary corrective actions. AI ethical dilemmas arise when the system is not well trained and hence leading to biased outputs. For example, if a machine relies on a specific indicator the result may be correct or completely wrong. According to Geirhos et al. "a deep neural network may appear to classify cows perfectly well but fails when tested on pictures where cows appear outside the typical grass landscape. In that case, 'grass' is the system's shortcut indicator for 'cow'" (Geirhos et al, 2020). In another case scenario, a job applicant may be selected over another due to historical discrimination over certain groups.

3.6 Privacy, Data Protection, Security (and Cybersecurity), Surveillance

In the age of digitalization and AI dominance, it seems that the privacy of individuals is increasingly violated due to the big influence of technology on users. This violation often comes in the form of dependence, since many users are forced to expose their personal data in order to access specific services (Müller, V, 2020). This data collection creates a general understanding of who this person is. As a result, the user loses control over his/her personal data and the right to privacy. Also, privacy and AI is linked to government espionage, theft of credit and intellectual property, fraud and blackmail but on the other hand, is used for monitoring public and private spaces in terms of preventing delinquent behaviors in the society, the working environment and citizens through the purchase of General Data Protection Regulation (GDPR) systems (Nouri, S., 2020).

3.7 Control Over AI

Perhaps on the top of the ethical issues arising from the use of artificial intelligence is whether humans will be able to control AI. An international group of researchers warned of the potential risks of creating overly powerful and standalone software (Humans Won't Be Able to Control Artificial Intelligence, Scientists Warn, 2021). Specifically, control over AI affects public trust, which is mainly related to the concept of "superintelligence", meaning that as artificial intelligence evolves to the point that it exceeds human abilities, it may reach the point where it takes control over our resources and

outcompetes our species, eventually leading to human extinction (Bird et al., 2020). An associated fear is that, even if an AI system was developed in such a way that its goals were fully aligned with human needs, it could develop for itself unexpected sub-goals that are not (Bird et al., 2020). Despite the fact that the majority of researchers agree that this is far from reality to take place, hence it is of major significance to maintain trust in AI through the ultimate human's oversight over this technology.

3.8 Misinformation, Fake News and Fraud

Fake news have already created high levels of distrust towards mass media, established institutions all over the world and politics (Cassauwers, 2019). New technological advancements, like AI gradually may worsen the situation, launching a new era of fake news and online misinformation. Based on Francesco Nucci, "Fake news is not a mathematical question of algorithms and data, but a very philosophical question of how we deal with the truth" (Cassauwers, 2019). Definitely, it is not an issue of detection, but mainly it is a problem of trust and also it constitutes a lack of critical thinking, which cannot be solved through technology. In addition, the use of AI for fraud prevention is not something new. In fact, AI solutions can be used to promote security across many business sectors as well as "sensitive" procedures such as the protection of voters' opinion during election campaigns.

3.9 Kids Ethics

Children interact with smart devises through toys, videos and adaptive-learning software which proposes with whom to play with, what to learn and watch. Apart from the positive effects of AI as aided education, health and diagnoses, societal aspects of inclusion or equity, AI is impacting children's privacy, safety and security, simply because children interact with systems not designed for them (UNICEF, 2020). Algorithms may be biased if they do not consider children's varied characteristics, if the input of data is not recorded in a systematic way or if they are grounded on statistical analysis of former cases and criteria obtained from varied databases, if there is lack of data protection and privacy rights, and when they do not count for access to technology and social inequalities. AI systems and policies shall be developed in a way to support children well-being, ensure their inclusion, and consider their data protection and privacy (UNICEF, 2020).

4. Conclusions

Hence, the critical issue that arises from the previous analysis is how to make AI more ethical. Law and data protection regulations could act as a tool to protect against AI discriminations. Regarding learning, one is certain, that machines don't learn as humans because intelligence is not only a relation of knowledge of facts, risks and opportunities, but the more data collected, the safer the result of the

judgment. For children, AI systems and policies shall be designed in a way to support their well-being, ensure their inclusion, and consider their data protection and privacy, as well as to account for the different socioeconomic, geographic setting. To deal with bias the key solution is diversity in data sets, design teams, participants, and validation process and as for Explainability and transparency issues, all data sets have to be commemorated with metadata specifying the source, intended use, user rights, etc., and the algorithms used must be able to be reproduced, easily traced and be fixed at any time.

In conclusion, addressing ethical dilemmas requires, above all, a holistic approach since it is not just a matter of technology and tools, but also about leadership, international standards and proper rules. Thus, it is considered essential a global alliance e.g., by setting up AI advisory bodies, to provide a universal regulatory framework on the ethical side of AI and commonly accepted global standards that will govern how humans interact with it.

References

Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Winfield, A. (2020). The ethics of artificial intelligence: Issues and initiatives. *European Parliamentary Research Service, 1*–128. https://data.europa.eu/doi/10.2861/6644

Bonnemains, V., Saurel, C., & Tessier, C. (2018). Embedded ethics: some technical and ethical challenges. *Ethics and Information Technology, 20*(1), 41-58.

Bossmann, J. (2016). *Top 9 ethical issues in artificial intelligence*. World Economic Forum. https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/

Bradshaw, S., & Howard, P. (2017). Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. In *Computational Propaganda Research Project* (pp. 1–37). Oxford Internet Institute.

Bryson, J. J. (2019). The Past Decade and Future of AI's Impact on Society. In *Towards a New Enlightenment? A Transcendent Decade* (150–185). Turner.

Burbidge, D., Briggs, A., & Reiss, M. J. (2020*). Citizenship in a Networked Age: An agenda for rebuilding our civic ideals*. University of Oxford.

Cassauwers, T. (2019). *Can artificial intelligence help end fake news?* Horizon: The EU Research & Innovation Magazine. https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news

Duhigg, C. (2012, February 16). *How companies learn your secrets*. New York Times. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Efthymiou-Egleton, I. P., (2020). Non-Cognitive Skills and AI: A New Era of Learning and Development. *Case Studies Journal, 9*(10), 42-46.

Efthymiou-Egleton, I. P., Egleton, T. W. E., & Sidiropoulos, S. (2020). Artificial Intelligence (AI) in Politics: Should Political AI be Controlled? *International Journal of Innovative Science and Research Technology, 5*(2).

Filippo, R., Hilligoss, H. Krishnamurthy, V. Bavitz, Ch., Kim, L. (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks*. The Berkman Klein Center for Internet & Society of Harvard University. https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights

Frey, C. B., & Osborne, M. (2013). *The future of employment: How susceptible are jobs to computerisation?*. Working Paper, Oxford Martin Programme on Technology and Employment. https://www.oxfordmartin.ox.ac.uk/publications/the-future-of-employment/

Humans Won't Be Able to Control Artificial Intelligence, Scientists Warn. (n.d.). Available from: https://www.entrepreneur.com/article/363284

Geirhos, R., Jacobsen, J. H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence, 2*(11), 665-673. Available at: https://arxiv.org/pdf/2004.07780.pdf

Gerards, J., & Zuiderveen Borgesius, F. (2020). Protected Grounds and the System of Non-Discrimination Law in the Context of Algorithmic Decision-Making and Artificial Intelligence. *Colorado Technology Law Journal, 20*. https://ctlj.colorado.edu/wp-content/uploads/2022/06/PROTECTED-GROUNDS-AND-THE-SYSTEM-OF-NON-DISCRIMINATION-LAW-IN-THE-CONTEXT-OF-ALGORITHMIC-DECISION-MAKING-AND-ARTIFICIAL-INTELLIGENCE.pdf

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15-25. https://doi.org/10.1016/j.bushor.2018.08.004

Koski, O. (2018). Work in the age of artificial intelligence: Four perspectives on the economy, employment, skills and ethics. Available at: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf

Mark, R. (2019). Ethics of using AI and big data in agriculture: The case of a large agriculture multinational. *The ORBIT Journal, 2*(2), 1-27.

Müller, V. (2020). *Ethics of Artificial Intelligence and Robotics*. Center for the study of Language and information of Stanford. https://plato.stanford.edu/entries/ethics-ai/

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A, 376*(2133). https://doi.org/10.1098/rsta.2018.0089

Nouri, S. (2020). *How AI Is Making An Impact On The Surveillance World*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2020/12/04/how-ai-is-making-an-impact-on-the-surveillance-world/

Ryan, M. (2019a). Ethics of using AI and big data in agriculture: the case of a large agriculture multinational. *ORBIT Journal, 2*(2). https://doi.org/10.29297/orbit.v2i2.109

Ryan, M. (2019b). Ethics of Public Use of AI and Big Data: The Case of Amsterdam's Crowdedness Project. *The ORBIT Journal, 2*(2). https://doi.org/10.29297/orbit.v2i1.101

Ryan, M. (2020). The future of transportation: ethical, legal, social and economic impacts of self-driving vehicles in the year 2025. *Science and engineering ethics, 26*(3), 1185-1208.

Ryan, M. and Gregory, A. (2019). Ethics of using smart city AI and big data: the case of four large European cities. *ORBIT Journal, 2*(2). https://doi.org/10.29297/orbit.v2i2.110

Siegfried, T. (2020). *Why some artificial intelligence is smart until it's dumb*. Knowable Magazine | Annual Reviews. https://knowablemagazine.org/article/technology/2020/why-some-artificial-intelligence-smart-until-its-dumb

Silberg J., &Manyika, J. (2019). *Notes from the AI frontier: Tackling bias in AI (and in humans)*. McKinsey Global Institute. https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

Stix, C. (2019). *A Survey of the European Union's Artificial Intelligence Ecosystem*. Leverhulme Centre for the Future of Intelligence, Cambridge.

Torresen, J. (2018). A Review of Future and Ethical Perspectives of Robotics and AI. *Frontiers in Robotics and AI, 4*, 75. https://doi.org/10.3389/frobt.2017.00075

UNICEF. (2020). *Policy guidance on AI for children*. New York: United Nations Children's Fund. https://www.unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf

Walch, K. (2019). *Ethical Concerns of AI*. Forbes. https://www.forbes.com/sites/cognitiveworld/2020/12/29/ethical-concerns-of-ai/

Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics, 2*(2), 46-48. https://doi.org/10.1038/s41928-019-0213-6