The Greek Review of Social Research, 2024, 162
Print ISSN: 0013-9696 Online ISSN: 2241-8512

BOOK REVIEW

Završnik, A. and Simončič, K. (Eds.) (2023). *Artificial Intelligence, social harms and human rights*. Palgrave Macmillan (XIV+276 pages). ISBN: 978-3-031-19148-0. https://doi.org/10.1007/978-3-031-19149-7.

The edited volume under examination is intended to fill a significant gap in the extant literature on the need for embedding AI developments in an appropriate legal and ethical framework, by suggesting fruitful ways to ensure that human rights violations are avoided. The book is divided into two separate parts: the first addresses human rights violations as a byproduct of AI systems in various domains (border control, surveillance, facial recognition); contributors in the second part proffer policy recommendations to address these major ethical challenges. The book drAWS on interdisciplinary research, appealing to scholars in a diversity of cognitive fields: AI ethics, political sociology, STS studies, computational ethics, criminology, and security studies.

In the first chapter entitled "Artificial Intelligence and Sentencing from a Human Rights Perspective", the authors comment on whether the functional capability of the criminal justice system could benefit from an algorithmic boost of efficiency. They thus advocate a robust legal culture centred on improving sentencing practice through deliberative processes, rather than replacing legal practice with technology solutions by which human judgment appears to be intrinsically discredited. In the second chapter, Patricia Faraldo Cabana focuses on the legal challenges that accompany the use of automated facial recognition technologies for law enforcement and forensic purposes. Albeit these technologies are expected to serve as a mechanism that addresses the perceived need for improved security, there remain ethical concerns that the absence of sufficient regulations endangers fundamental rights to human dignity and privacy.

Kristian Humble in the third chapter explores the extent to which the international community, states and corporations exhibit due concern

Received: August 18, 2023 Accepted: October 16, 2023 Published: January 31, 2024

about the implications of developing fully autonomous weapons systems (AWS), irrespective of any human input. These systems are inherently capable of making autonomous decisions in times of conflict, thus reflecting an ongoing process of militarization of AI. The ethical concerns of employing automated weaponry in modern warfare will be aggravated "as technological advances start to enable AWS with cognitive human-like decision-making" (p.72).

In the fourth chapter, Karmen Lutman identifies controversies centred on discriminatory practices arising in areas of private law in which algorithms are used in decision-making processes, such as those related to loan financing, marketing, personnel recruitment and employment, and insurance. The EU has advanced a legal framework that prohibits discriminatory practices of AI in horizontal relationships between private individuals and more specifically, in employment matters and access to, and the supply of, goods and services. The author convincingly argues that "protection against algorithmic pricing is insufficient since it protects consumers only against unequal treatment based on sex, race, and ethnic origin. This can lead to discrimination on other grounds that deserve protection, for example, sexual orientation, and promotes consumer exploitation and social sorting" (p. 94).

In the fifth chapter, Aleš Završnik emphasizes the interaction between ethics and law in a case study of legal and ethical assessments of access, collection, and multiple types of processing of personal data for computer vision. A legal perspective on fairness in AI comprises among others, a normative framework on the prohibition of discrimination, personal data protection law and the protection of intellectual property rights. The author argues that both ethical and legal frameworks operate jointly in AI governance, thus mitigating the negative societal effects of AI systems.

Mariavittoria Catanzariti in the sixth chapter explores the reasons that motivated the EU to develop an ethical approach to AI, seeking to examine the degree to which the ethical principles for a trustworthy AI should be based on compliance with fundamental rights. The author contends that neither an ethical approach nor a mere legal design of AI systems can effectively address the challenges of algorithmic inferences that affect both individuals and society. The author concludes that "In substance, the ethical approach, strongly encouraged by the European Parliament, relates to the legal conceptualization of the threshold of acceptability of AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights" (pp. 156-157).

In the next chapter, Ljupčo Todorovski introduces computational ethics, a field of AI that involves algorithms for undertaking ethically permissible decisions. The author epitomizes typical, exemplary approaches to these emerging issues and assesses them in the light of their capability to properly incorporate specific ethical principles and theories. In sum, "computational complexity of decision problems related to ethical decision-making is often prohibitive to implementation of efficient algorithms" (p. 175).

Lottie Lane in the eighth chapter exemplifies the negative effect that AI systems produced by private companies induce on human rights, for instance through inaugurating discriminatory access to goods and services. The author proceeds to underscore the need for a general legal framework on business and human rights and advocates corporate responsibility initiatives that fully respect human rights, especially regarding ethical challenges raised by recent AI developments. What is needed is the "collaboration between practitioners, law policymakers and academics to clarify AI businesses' responsibilities and to translate corporate respect for human rights vis-a-vis AI into workable processes in practice" (p. 200).

In chapter nine, Iva Ramuš Cvetkovič and Marko Drobnjak discuss the use of international space law as an inspiration model for terrestrial AI regulation because of maximizing harm prevention. The authors posit that certain space law provisions can inform a comprehensive binding framework aimed at preventing harm both on Earth (for example, predicting natural disasters) and in outer space (monitoring environmental harm through data collection), as well as by fairly distributing social benefits. This endeavour necessitates ethical principles that are deemed relevant to the prevention of harm, namely, *transparency*, *non-maleficence*, *responsibility and accountability*, *beneficence, sustainability,* and *solidarity* (p. 218). However, the deployment of these advanced technologies has not yet realized its full potential, thus it remains unclear which state or private actor will first initiate fully autonomous, strong AI systems, and how such AI will be used (p. 230).

Last but not least, in the final chapter Katja Simončič and Tonja Jerele underscore the need for democratizing the governance of AI by enacting a substantial shift, from big tech monopolies to cooperatives. They thus indicate the socially detrimental aspects of AI and advocate pathways toward an AI devoted to promoting the common good. This ideal necessitates the construction of cooperatives as a model for democratized, people-centred big tech companies based on seven normative principles that may outweigh the rampant commercialization of science and

technology: *voluntary and open membership*, *democratic member control, member economic participation, autonomy and independence, education/ training and information, cooperation among cooperatives and finally, concern for the community* (pp.252-253). Proponents of distributed, decentralized, and democratized development of AI hold the view that marginalized and vulnerable groups are expected to significantly benefit from this substantial power shift, from a few monopolistic corporations to a cluster of cooperatives (pp. 258-259).

These contributions provide permeating insights into the need for employing a functional normative framework capable of regulating AI systems, yet in some instances, the overall picture seems to be more complicated. Despite the prevalence of certain AI ethical considerations, most corporations remain so far strategically unprepared to effectively respond to persistent public concerns, thus exposing themselves to the peril of ethical failure. AI development operates as a 'double-edged sword' that involves not only unpredictable but frequently unintended consequences. For instance, AI could support meaningful work through enhanced human learning, skills, and competencies; yet, at the same time, AI could diminish meaningful work through various pathways, namely: through reduced human role in the work process, through weaker feelings of belongingness and less human interaction, through lower levels of personal autonomy, as well as through unfair distribution of benefits and burdens (Bankins & Formosa, 2023).

Not unexpectedly, the book highlights the need for elaborating a coherent framework intended to encompass the normative principles employed in the attempt to regulate AI development. In the absence of adequate regulation in the AI domain, we could witness human rights infringement and potential violation of social norms, reflecting a state in which public agencies would have practically relinquished their ethical obligation to meaningfully protect and ensure a bundle of inalienable rights. The pervasiveness of AI in a unique private sphere tends to erode the moral autonomy of the individual, thus colonizing personal lifeworld in a non-conventional, unexpected manner. Unprecedented challenges for policymakers arise as machines acquire the ability to learn and evolve, thus becoming autonomous in their decision-making. If AI assumes the ability to act *independently of human intervention* based on its self-awareness, it will irrevocably enjoy the status of an autonomous moral entity.

The book under consideration explicates precisely how such AI systems become increasingly ubiquitous in social life. Accordingly, policymakers

assume a pivotal role in ensuring that AI-related decisions are both just and socially beneficial. Undoubtedly, it is ultimately the human moral compass that we are relying on to employ these powerful technologies in other than a socially detrimental way (De Cremer & Narayanan, 2023). For instance, there is a need for elaborating on an ethical framework for designing AI systems in medical practice, based on the core pillars of responsible design: transparency, fairness and justice, safety and wellbeing, accountability, and collaboration. Interestingly, Nikolinakos (2023) identifies five primary ethical imperatives and their correlated values underlying the development, deployment, and use of AI systems, namely: *respect for human autonomy*, *prevention of harm (non-maleficence)*, *fairness/justice*, *explicability* and finally *beneficence,* as the core principle of shaping AI technologies potentially beneficial to humanity.

To summarize, this important volume articulates a set of valuable policy recommendations. Ethical technology design and implementation is expected to help policymakers, software developers and academic researchers in seeking effective solutions to address new, unpredictable challenges in a rapidly changing societal environment. We thus deem it necessary to argue in favour of appropriate discursive strategies that are in a position to promote the responsible and ethical development of generative AI (Cheng and Liu, 2023). Such thematic intertextuality between ethical and legal discourses facilitates processes of convergence of narrative-ideological structures, which in turn shape ethical frameworks that enable a holistic approach to the prevailing human-AI interaction paradigm. These emerging technologies should be deployed in conjunction with societal values and within the boundaries of fostering human empowerment, otherwise humans would be exposed to various perils, namely: reduced human control, repudiation of human responsibility and more importantly, an ongoing process of eroding human self-determination (Paraman & Anamalah, 2023). Whether in a possible dystopian scenario, these emerging posthuman agents will subdue biological humanity, or AI and humanity might join in a collaborative endeavour to create a digital utopian society and generate new dimensions of rationality, remains an issue of ultimate choice, entrenched in the sphere of human freedom.

Prof. George Gotsis
Department of the History and Philosophy of Science
National and Kapodistrian University of Athens

REFERENCES

Bankins, S. and Formosa, P. (2023). The ethical implications of Artificial Intelligence (AI) for meaningful work. *Journal of Business Ethics*, 185 (4), pp.725-740.

Cheng, L. and Liu, X. (2023). From principles to practices: The intertextual interaction between AI ethical and legal discourses. *International Journal of Legal Discourse*, 8 (1), pp.31-52.

De Cremer, D. and Narayanan, D. (2023). How AI tools can-and cannot-help organizations become more ethical. *Frontiers in Artificial Intelligence*, 6, 1093712.

Nikolinakos, N.T. (2023). Ethical Principles for Trustworthy AI. In *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies* (pp. 101-166). Cham: Springer.

Paraman, P. and Anamalah, S. (2023). Ethical artificial intelligence framework for a good AI society: Principles, opportunities, and perils. *AI and Society*, 38 (2), pp. 595-611.

This paper is an integral part of EU ERASMUS+ Project ETHICS4CHALLENGES (E4C): *Innovative Ethics Education for Major Technological and Scientific Challenges*.