AI & SOCIETY
https://doi.org/10.1007/s00146-024-01964-6

EDITORIAL

Machine theology or artificial sainthood!

Karamjit S. Gill[1]

Accepted: 3 May 2024

As we explore ethical impacts and implication of Open AI (ChatGPT, LLMS), we may ponder over the question: If AI provides huge potential of benefits to society, ranging from agriculture, medicine, health care, education and learning, workplace innovation, and poverty elimination, then

> Then why are we now facing a cultural crisis of embedding AI in societies?

In other words, what is it about the engineering of predictive algorithms which drive social platforms that shapes our concerns about conspiracy theories, security, privacy, prejudice, and identity crisis. These concerns raise the question: how do we mould socially responsible AI tools for societal benefits whilst mitigating the danger of falling into the trap of Faustian seduction in which we bargain our soul with the machine in exchange for getting what we desire. During the 1980s, the challenge and concern was how to mitigate the consequence of turning 'judgment' to 'calculation', and now, the 2020s challenge and concern is about the impact and implication of turning the human (outer-inner) into data. In pursuing this exploration, we are reminded by Mihai Nadin (2019) of the 'deterministic theology' of computational algorithms that only mimics creative aspects without being creative, and mimics meanings without being meaningful. He asks us to keep in mind Dreyfus's concern about the context-free processing of data as a path towards emulating intelligence. He says that the idea of synthesised intelligence mimicking human intelligence, for example winning a game (chess or any other), misses the most important aspect: the creation of the game itself, as one of many instances in which human beings shape their own condition. The game of chess documents learning, the ability to represent and to make associations, the understanding of reward, and the awareness of aesthetic expression. He further elaborates that the notion of computation has changed the world more than any previous expressions of knowledge. However, in the particular algorithmic embodiment of know-how, computation is closed to meaning. Therefore, computer-based data processing can only mimic life's creative aspects, without being creative itself. AI's current record of accomplishments shows that it automates tasks associated with machine intelligence, without being intelligent itself. Mistaking the abstract (computation) for the concrete (computer) has led to the religion of "everything is an output of computation"—even the humankind that conceived the computer. The hypostatized role of computers explains the increased dependence on them. The convergence machine called *deep learning* is only the most recent form through which the deterministic theology of the machine claims more than what it actually is: extremely effective data processing. Nadin (ibid.) argues that a proper understanding of complexity, as well as the need to distinguish between the reactive nature of the artificial and the anticipatory nature of the living are suggested as practical responses to the challenges posed by machine theology, and thus, we need to be mindful of falling into the trap of an artificial sainthood of the ethical machine. Whilst we have been reflecting on the transition from the Turing imitation game to machine mythology, we are invited to meet AI powered digital clones, aspiring to transmit wisdom to their virtual disciples. It is as if by communicating the 'personality, thoughts, and philosophy' of their makers, these artificial saints can contemplate away human condition, say, ecstasy, pain, and suffering, as they walk and talk virtually.

Our authors in this volume contribute to theoretical debates on ethics regarding the digital driven institutions, public administration, public policy and services, raising and articulating issues of trust, ethics of well-being transparency, accountability, values, identity, moral constructs, and judgement. For example, in commenting on the increasing demand for transparency in AI, the article, 'AI and the expert; a blueprint for the ethical use of opaque AI' (this volume) observes the debate on "epistemic double standards", whether the standards for transparency in AI ought to be higher than, or equivalent to, our standards for ordinary human reasoners. The article suggests that a more fruitful

✉ Karamjit S. Gill
editoraisoc@yahoo.co.uk

[1] University of Brighton, Brighton, UK

Published online: 10 May 2024

exploration of this question will involve a different comparison class, for example how highly trained experts routinely make judgments whilst not being able to provide an explanation of the reasoning behind these judgments. From this perspective, the article suggests that it is better to reframe our question regarding the appropriate standards of transparency in AI to one that asks when, why, and to what degree it would be ethical to accept opacity in AI.

Delving into the essential role of practical wisdom rooted in ethos of 'purpose', the article, 'Apprehending AI moral purpose in practical wisdom' (this volume) argues for grounding an AI system's apprehension of reality in a sociotechnical moral process that is committed to orienting AI development and action in light of a pluralistic, diverse interpretation of what is good for society. The suggestion is that these apprehensions are structured by moral and other social-cognitive schemas that mediate moral constructs, including individual commitments and other motivating identifications, in a process of informing historical reality. It further argues that defining AI purpose in terms of the human struggle to determine the moral "ought" (versus the way reality is) can overcome practical and normative limitations of principle-based approaches to AI ethics without requiring idiosyncratic moral reasoning or a premature commitment to a single ethical theory, which could hinder navigating the moral implications of future unknowns.

Commenting upon the ethical discourse on autonomous vehicles, the article 'Bowling alone in the autonomous vehicle: the ethics of well-being in the driverless car' (this volume) offers a human well-being focus as an alternative to the behaviour of the vehicle-oriented focus in accident scenarios. This paper offers a different ethical prism: the implications of the autonomous vehicle for human well-being. As such, it contributes to the growing discourse on the wider societal and moral implications of the autonomous vehicle. It is asserted that human well-being focus of ethics is premised on the neo-Aristotelian approach which holds that as human beings, our well-being depends on developing and exercising our innate human capacities: to know, understand, love, be sociable, imagine, create and use our bodies, and use our willpower.

Commenting upon the ethical and legal implications of the claim that AI might become part of the person, the article, 'Might artificial intelligence become part of the person, and what are the key ethical and legal implications?', argues that this claim is based on the idea that AI tools may act in the same way as arms, hearts, or mental capacities are. The article further explores the consequences of this "*empersonification'*, and thus notes three practical implications: (1) AI-devices cease to exist as independent legal entities and come to enjoy the special legal protection of persons; (2) therefore, third parties such as manufacturers or authors of software lose (intellectual) property rights in device and software; (3) persons become responsible for the outputs of the empersonified AI-devices to the same degree that they are for desires or intentions arising from the depths of their unconscious. It is posited that this '*empersonification*' marks a new step in the long history of human–machine interaction that deserves critical ethical reflection and calls for a stronger value-aligned development of these technologies.

In pursuing the notion of social trust in digitally driven public administration, public policy, and services by societies, the article, 'Social trust and public digitalization' (this volume), argues that high-trust countries cannot do without the rule of law (formal rules of the game) and some institutionalised mechanisms of control and accountability. It notes, however, that the challenge is how to respond when rule-based overcontrol may crowd out trust, thus undermining the balance of trust between citizens and government, even in high-trust countries.

The article, 'AI ethics with Chinese characteristic' (this volume), sheds light on the way Chinese academia is shaping the national landscape of discussion on AI ethics. It is posited that from a short-term perspective, Chinese scholars concerns over AI ethics predominantly resemble the content of international ethical guidelines. However, from a long-term perspective, there is a recognition of the need for addressing these issues within cultural contexts. We are informed that amongst a wide range of solution proposals, Chinese scholars seem to prefer strong-binding regulations to those weak ethical guidelines.

Commenting on the self-regulation regimes that focus on softer ethics-based approaches, the article, 'The limitation of ethics-based approaches to regulating artificial intelligence' (this volume), explores the inherently uncertain effects of AI technologies on society, both short and long term. The article posits that it for this reason of uncertainty, many governments avoid strict command and control regulations, and instead rely on softer ethics-based approaches. Citing the example of the big tech-driven regulatory regime in Russia, the article notes how big tech companies avoid regulatory oversight by washing out concrete regulatory measures from the policy, thus regard unenforceable ethics-based self-regulation as a '*regulatory gift*'. However, the article notes that the '*gift*' can also unintentionally undermine the public interest by providing an opportunity for *ethics washing*.

Some of the key issues we face are: How do we design AI tools that align with societal goals such as those of trust and trustworthiness? Can the idea of the ethical machine drive towards taming ethical and moral dilemmas of alignment, beyond the Cartesian belief in human–AI co-evolution. Could a human–machine symbiotic framework as in Buber's conception of I-IT and I-Thou symbiosis and Shiva's Dance of harmony provide an alternative to the Faustian bargain? We then ask whether present day followers of the Cartesian faith allow themselves to heed Weizenbaum's warning that

humans and computers belong to separate and incommensurable realms (Gill 2024).

Data availability Not applicable.

References

Gill KS (2024) Eliza! A reckoning with Cartesian magic. AI Soc 39:1–3. https://doi.org/10.1007/s00146-024-01868

Nadin M (2019) Machine intelligence: a chimera. AI Soc 34:215–242. https://doi.org/10.1007/s00146-018-0842-8