

CAN AI BE ETHICAL AND MORAL?

Relevant for: Science & Technology | Topic: Robotics & Artificial Intelligence

To enjoy additional benefits

CONNECT WITH US

August 24, 2023 02:12 am | Updated 02:24 am IST

COMMENTS

SHARE

READ LATER

Can machines or AI be moral agents, or artificial moral agents (AMAs)? A wide body of literature suggests that machines can, “in some sense, be ethical agents responsible for their actions,” or “autonomous moral agents”. [File](#) | [Photo Credit: Reuters](#)

Increasingly, machines and Artificial Intelligence (AI) are assisting humans in decision-making, particularly in governance. Several countries are introducing AI regulations. Government agencies and policymakers are leveraging AI-powered tools to analyse complex patterns, forecast future scenarios, and provide more informed recommendations.

However, the adoption of AI in decision-making is not without its potential pitfalls. The biases inherent in AI, often a reflection of the biases in the data they are trained on or the perspectives of their developers, can lead to skewed or unjust outcomes and represent a significant challenge in the integration of AI into governance. Biases in AI warrant a whole other article. This article is about ethics in AI. Can AI be ethical and moral?

Immanuel Kant’s ethical philosophy emphasises autonomy, rationality, and the moral duty of individuals. Applying Kantian ethics to the use of AI in decision-making within governance could lead to serious concerns. If decisions that were once the purview of humans are delegated to algorithms, it could threaten the capacity for moral reasoning. The person or institution using AI could be considered to be abdicating their moral responsibility. This argument finds echoes in Isaac Asimov’s ‘Three Laws of Robotics’. The laws were designed to govern robotic behaviour, aiming for ethical actions, but within Asimov’s fictional world, the laws lead to unexpected and often paradoxical outcomes. The attempts to codify ethics into rules, whether for robots or complex AI-driven governmental decision-making, reveal the inherent challenges in translating human moral complexity into algorithmic form. The intertwining of Kant’s insistence on rational moral agency with Asimov’s fictional exploration of coded ethics illustrates the ethical challenges in delegating human functions to artificial entities.

Despite this, it is inevitable that AI would be used in governance decisions. In some countries, decision-making algorithms are even being used to determine the beneficiaries of social sector schemes. However, how can governments guarantee that decisions assisted by machines remain ethical and moral?

Ethics are complex. Programming ethics into a machine and AI is even more complex. As James Moore of the Dartmouth College said, “Programming a computer to be ethical is much more difficult than programming a computer to play world-champion chess... Chess is a simple

domain with well-defined legal moves. Ethics operates in a complex domain with some ill-defined legal moves.”

But can machines or AI be moral agents, or artificial moral agents (AMAs)? A wide body of literature suggests that machines can, “in some sense, be ethical agents responsible for their actions,” or “autonomous moral agents”. In Moore’s 2006 classification, four categories of machine agents relating to ethics are defined. The first is ethical impact agents: machines with ethical consequences, like robot jockeys, which don’t make ethical decisions but pose ethical considerations, such as altering the sport’s dynamics. The second is implicit ethical agents: machines with embedded safety or ethical guidelines, such as a safe autopilot system in planes, which follow set rules without actively deciding what is ethical. The third is explicit ethical agents which go beyond set rules, using formal methods to estimate the ethical value of options, like systems that balance financial investments with social responsibility. The fourth is full ethical agents which are capable of making and justifying ethical judgments, including reasonable explanations. An adult human is a full ethical agent, and so would be an advanced AI with a similar understanding of ethics.

But is it that easy to create AMAs, especially the third and fourth? No. A peer-reviewed paper published in *Science and Engineering Ethics* found that from a technological standpoint, artificial agents are still far from being able to replace human judgment in complex, unpredictable, or unclear ethical scenarios. There is another issue, i.e., of bounded ethicality. Hagendorf and Danks (2022) fed prompts to Delphi, a research prototype designed to model people’s moral judgments. They found that similar to humans, machines like Delphi may also engage in immoral behaviour if framed in a way that detaches ethical principles from the act itself. This suggests that human patterns of moral disengagement could translate into machine-bounded ethicality. Moral disengagement is a key aspect of bounded ethical decision-making, allowing people to act against their ethics without guilt through techniques like moral justifications.

Today, many machine predictions are deployed to assist in decisions where a human decision-maker retains the ultimate decision-making authority. Eventually, governments would delegate a few rudimentary decisions to the machines. But what if the decision taken on the basis of machine predictions or by a machine itself is immoral or unethical? Who will be held responsible? Will the liability fall on the AI system, the developer of the AI, or the official who relied on the AI’s data? The notion of punishing the AI system becomes problematic, as it lacks the ability to experience suffering or bear guilt. How would the developer or the official then be made accountable? These are the questions which governments would grapple with eventually. Programming ethics into machines is complex, and the world must proceed cautiously.

Aditya Sinha is OSD (Research), Economic Advisory Council to the Prime Minister. Tweets @adityasinha004.

COMMENTS

SHARE

[Artificial Intelligence](#) / [ethics](#)

BACK TO TOP

Comments have to be in English, and in full sentences. They cannot be abusive or personal. Please abide by our [community guidelines](#) for posting your comments.

We have migrated to a new commenting platform. If you are already a registered user of The

Hindu and logged in, you may continue to engage with our articles. If you do not have an account please register and login to post comments. Users can access their older comments by logging into their accounts on Vuukle.

END

Downloaded from **crackIAS.com**

© **Zuccess App** by crackIAS.com

CrackIAS.com