

Artificial Intelligence and Artificial Moral Agents

Alvin Chen, Academia Sinica

Abstract :

With the recent development of generative artificial intelligence, questions like “Can AI be responsible of its wrongs?” and “if can, how?” become increasingly pressing; the so-called “responsibility gap”. These questions are mostly framed in terms familiar to the field of machine ethics, a field of studies emphasizing ethical reflections on the development of machines. But conceptions of artificial agent with moral capacity in legal and political theory predates the emergence of machine ethics and the development of generative AI. This paper argues that the so-called “responsibility gap” exists only because of the account of artificial agent in machine ethics. It shows that, once insights from legal and political philosophy are taken, there will be a much clearer understanding of the characters of artificial moral agents and a less troubling undertaking of the responsibility of AI. In terms of specifics, the paper addresses the following tripartite question: Is AI an artificial moral agent? If it is, what kind of artificial moral agent is it? Why does it matter for AI to be an artificial moral agent? An answer to the third part of this question quintessentially presents a response to the “responsibility gap” by reassessing how AI can take responsibility. The paper concludes with reflections on its practical implications, notably to AI in medicine.

Keywords: Artificial Intelligence (AI), Artificial Moral Agents (AMAs), responsibility gap, rational belief, relational responsibility