

Curbing the AI Calamities: a call for philosophical and Strategic shift

Fareed Alghayath

Kapodistrian University of Athens

Abstract

The recent coverage of Google's Lambda-AI's alleged conscious capacities has flared the discussion pertaining to AI ethics and governance. Thus, the containment or limitation of potential AI Harms has become both in the popular press and in the academic context a more pressing matter. Most of the existing literature focuses on specifying either a set of principles that AI developers ought to adopt, or a list of formalized regulations that will control AI development. However, shifting from *Is* to *Ought* is neither inevitable nor a straight path; there are several grounds for being skeptical about the epistemological legitimacy of drawing normative conclusions from (supposedly) factual premises. Likewise, constructing regulatory systems for a rapidly changing industry is extremely difficult to maintain, especially if it is engendered outside a convincing moral paradigm or a well-grounded philosophical perspective. Accordingly, the paper articulates a different approach towards curbing potential AI calamities. Our approach is a combination of (i) a paradigm shift on how we interpret AI, the prospects of its rationality, and certain principles of AI ethics, on the one hand, and on the other hand, (ii) a recommendation to adopt a well-established framework of corporate risk management that is used for curbing the undesired consequences of untested innovation, in general. As it will be elaborated, the suggested approach may complement some of the currently operating approaches, and offer a more maintainable strategy for dealing with potential AI harms, taking into consideration both the epistemological debate over the projected development of AI, and its actual on ground operation.

Keywords: AI Philosophy, Risk Management, Ethics, Governance, Strategy