

Uncovering the Gap: Challenging the Agential Nature of AI Responsibility Problems (Extended Abstract)

Joan Llorca Albareda

Department of Philosophy I, University of Granada, Spain
joanllorca@ugr.es

Extended Abstract

In this paper, I will argue that the responsibility gap arising from new AI systems is reducible to the problem of many hands and collective agency. Systematic analysis of the agential dimension of AI will lead me to outline a disjunctive between the two problems. Either we reduce individual responsibility gaps to the many hands, or we abandon the individual dimension and accept the possibility of responsible collective agencies. Moreover, I will adduce that this conclusion reveals an underlying weakness in the various discussions of AI ethics: the habitual lack of attention to the question of the disciplinary boundaries of AI ethics. This absence has made it difficult to identify the specifics of the responsibility gap arising from new AI systems as compared to the responsibility gaps of other applied ethics. Lastly, I will be concerned with outlining these specific aspects.

The argumentation will proceed as follows. First, I will develop the general concept of the responsibility gap by attending to both the elements and conditions of responsibility. This will reveal that the responsibility gap arising from the new AI systems depends on the breach of the agential control condition. Subsequently, I will present the three strategies that have been pursued to address the responsibility gap. Some authors have argued that the concept of responsibility is much richer and more dynamic than this interpretation suggests (Tigard 2020). Another group of authors have argued that the responsibility gap is inconsistent or of little importance (Hindriks and Veluwenkamp 2023). And, lastly, some authors have offered novel conceptual solutions to address the responsibility gap (Köhler 2020; Nyholm 2018). While each of them has strengths, I will defend that their weaknesses lead us to conduct a systematic analysis of the agential dimension of new AI systems.

Second, I will analyze three types of conceptions of AI agency: moral agency with responsibility, moral agency without responsibility, and neither moral agency nor

responsibility. I will defend that the first one leads us to understand AI from the concept of responsible collective agency and the last two are reducible to the many hands problem. Finally, I will expose the lack of attention devoted to the disciplinary boundaries of AI ethics and argue that the problem of ethical proliferation (Sætra and Danaher 2022) is at the root of the difficulties that the responsibility gap of new AI systems has generated. After that, I will show the specific aspects of the responsibility gap produced by these systems. From an internal point of view, it creates new objects of responsibility: their large and complex databases, and the criteria used in algorithmic training. From an external point of view, it has greater impact than other technologies: the increase in the number of systems and their ubiquity.

Keywords: responsibility gap, control condition, agential, ethical proliferation, many hands, collective agency

References

- Hindriks, F.; and Veluwenkamp, H. 2023. The risks of autonomous machines: from responsibility gaps to control gaps. *Synthese* 201(1): 21. <https://doi.org/10.1007/s11229-022-04001-5>
- Köhler, S. 2020. Instrumental robots. *Science and Engineering Ethics* 26(6): 3121-3141. <https://doi.org/10.1007/s11948-020-00259-5>
- Nyholm, S. 2018. Attributing agency to automated systems: Reflections on human-robot collaborations and responsibility-loci. *Science and Engineering Ethics* 24(4): 1201-1219. <https://doi.org/10.1007/s11948-017-9943-x>
- Sætra, H. S.; and Danaher, J. 2022. To Each Technology Its Own Ethics: The Problem of Ethical Proliferation. *Philosophy & Technology* 35(4): 1-26. <https://doi.org/10.1007/s13347-022-00591-7>
- Tigard, D. W. 2020. There is no techno-responsibility gap. *Philosophy & Technology* 34(3): 589-607. <https://doi.org/10.1007/s13347-020-00414-7>