Computational Neuroethics

In this paper, I delineate some core aspects of the nascent field of computational ethics and stress its relevance for neuroethics. I suggest that neuroethics, the ethics of AI, and computational approaches to moral psychology should be reciprocally more integrated and highlight emerging issues and lines of research of great scientific, public, and philosophical relevance. This field – computational neuroethics – overlaps with traditional issues in neuroethics and the ethics of AI, but there are reasons for considering it a specific field with peculiar features.

On the one hand, AI ethics – i.e. developing ethical and regulatory frameworks for AI, and implementing moral decision-making in machines – is now an established multidisciplinary field of inquiry and application. It encompasses issues such as algorithmic accountability and transparency (understanding, explaining and justifying AI decisions), biases, fairness, and the social impact of AI. However, these issues are often approached without dealing with the neurocognitive sciences: how can we improve AI systems in light of our knowledge of the human brain, mind, and behavior? How do/could these latter interact with AI systems? On the other hand, neuroethics, moral psychology, and ethics more generally are starting to appreciate the contribution of computational tools and AI for coping with moral problems, and even understanding and improving our conceptions of ethics.

Drawing from a paradigmatic conceptualization by Roskies (2002), I suggest conceiving this emerging research field as including (1) the ethics of computational neuroscience and (2) the computational (neuro)science of ethics. Even more than in traditional neuroethics, these two perspectives are and should be treated as deeply interdependent.

References

Awad, E., et al. (2022). Computational ethics. Trends in Cognitive Sciences, 26(5), 388-405.

Etzioni, A., & Etzioni, O. (2016). AI assisted ethics. Ethics and Information Technology, 18, 149-156.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. Ethics, 124(4), 695-726.

Ienca, M. (2019). Neuroethics meets Artificial Intelligence. The Neuroethics Blog [online]. http://www.theneuroethicsblog.com/2019/10/neuroethics-meets-artificial.html

Kleiman-Weiner, M., Saxe, R., & Tenenbaum, J. B. (2017). Learning a commonsense moral theory. Cognition, 167, 107-123.

Railton, P. (2014). The affective dog and its rational tale: Intuition and attunement. Ethics, 124(4), 813-859.

Rodríguez-López, B., & Rueda, J. (2023). Artificial moral experts: asking for ethical advice to artificial intelligent assistants. AI and Ethics, 1-9.

Roskies, A. (2002). Neuroethics for the new millenium. Neuron, 35(1), 21-23.

Sinnott-Armstrong, W., & Skorburg, J. A. (2021). How AI can AID bioethics. Journal of Practical Ethics, 9(1).

Vallor, S. (2016). Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press.