The meta-problem of artificial consciousness and moral status

A great deal of attention has recently been devoted to what can be referred to as the problem of artificial consciousness — that is, whether and under what conditions AI systems can be conscious (Koch, Tononi, 2008; Dehaene, Lau, Kouider, 2017; Hildt, 2022). A different approach is adopted here, and the meta-problem of artificial consciousness is formulated and discussed: why are we taking seriously the possibility of conscious AI? First, I show that the meta-problem of artificial consciousness is distinct and logically prior to the problem of artificial consciousness, and therefore should be given appropriate attention. Possible answers to the problem are then considered by taking into account three distinct sources of concern about conscious AI: artificial self-reports, analogy with biological organisms, and the systematic use of cognitive notions in the explanation of AI systems' behaviour. I show that none of these is sufficient for grounding discussions on artificial consciousness. On the contrary, they turn out to be misleading and end up polarizing the debate. I proceed by arguing that a more convincing reason for being serious about the possibility of conscious AI is that the functioning of some AI systems depends on recurrent processing, which is typically deemed necessary - and sometimes sufficient - for consciousness in biological organisms (Northoff, Lamme, 2020). That being said, I contend that the meta-problem of artificial consciousness is still largely open. I conclude by focusing on the ethical implications of the meta-problem, arguing that it triggers a problem of moral meta-uncertainty. As a matter of fact, not only are we uncertain about the possibility of conscious and therefore intrinsically valuable AI systems, but we also ignore whether the very question of artificial systems' intrinsic value is worth exploring.

- Dehaene, S., Lau, H., & Kouider, S. (2021). What is consciousness, and could machines have it? Robotics, AI, and Humanity: Science, Ethics, and Policy, 43-56.
- Hildt, E. (2022). The Prospects of Artificial Consciousness: Ethical Dimensions and Concerns. AJOB neuroscience.
- Koch, C., & Tononi, G. (2008). Can machines be conscious? Ieee Spectrum, 45(6), 55-59.
- Northoff, G., & Lamme, V. (2020). Neural signs and mechanisms of consciousness: is there a potential convergence of theories of consciousness in sight? Neuroscience & Biobehavioral Reviews, 118, 568-587.