

# Review of: "Questioning the Moratorium on Synthetic Phenomenology"

Carles Salazar<sup>1</sup>

<sup>1</sup> Universitat de Lleida

**Potential competing interests:** No potential competing interests to declare.

## Synthetic phenomenology and counterfeit humans

Krzanowski presents a compelling argument against Metzinger's proposal for a moratorium on the development of synthetic phenomenology. Metzinger contends that such a moratorium is necessary to avoid potential massive suffering inflicted upon conscious AI systems if research in this area continues unchecked. Krzanowski counters by asserting that Metzinger's proposal lacks substantiation, primarily because it rests on an assumed cognitive equivalence between conscious AI systems and human beings, a notion that is far from reality. Therefore, without this cognitive equivalence, there's no reason to assume that research on AI systems would cause harm, as they lack the capability to experience harm in any meaningful way.

The question of whether conscious AI systems could experience suffering is debatable. On one hand, there's no inherent reason to deny such a capacity, as any form of consciousness should theoretically include the ability to suffer. However, current AI research is nowhere near producing anything resembling synthetic phenomenology. Thus, while the possibility cannot be entirely dismissed for the future, it remains an entirely speculative concept at present. This, according to Krzanowski (if I understood him correctly), undermines the validity of Metzinger's proposal.

However, there's another significant reason supporting the idea of a moratorium, one which Krzanowski surprisingly overlooks. This reason concerns not the creation of conscious AI systems, but rather the development of AI systems that mimic conscious beings, a concept Dennett aptly terms 'counterfeiting humans'.<sup>[1]</sup> Just as counterfeiting money poses a threat to the economy, counterfeiting humans could jeopardize our current moral frameworks. It's alarming that Krzanowski fails to address this issue, as unlike Metzinger's speculations, it represents a tangible danger already emerging, albeit within a limited scope, with advancements such as the latest versions of LLM. Thus, while a moratorium on synthetic phenomenology research may not be entirely unreasonable, the rationale behind it differs from Metzinger's. It's not about preventing suffering in conscious AI systems, as they do not exist; rather, it's about preventing the confusion between genuine humans and artificial entities.

[1] [Counterfeiting Humans: A Conversation with Daniel Dennett and Susan Schneider \(youtube.com\)](https://www.youtube.com/watch?v=...)



<https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/>