

Artificial Consciousness: From Anthropomorphism and Metaphysical Speculation to Rigorous Scientific Research

Recent advances in AI have moved questions of AI consciousness from purely theoretical speculation to potentially practical concerns. The growing public discourse around AI capabilities and limitations makes it essential to develop rigorous approaches for evaluating claims about AI consciousness while acknowledging the role of social understanding in shaping these discussions.

Drawing on philosophical and AI ethics perspectives, this panel examines consciousness, AI, and their intersection by discussing the following questions:

1. Does consciousness require a biological basis, and, if yes, in which way?
2. Who will, or should, settle the question of whether AI is conscious, if at all: science or public discourse?

In discussing these questions, we will explore whether consciousness can be meaningfully conceived independently of biology, or whether biological foundations are inescapable. These investigations aim to advance both scientific and philosophical approaches to consciousness studies.

The panel intends to stimulate discussion on how to develop approaches to understanding consciousness that can accommodate both accounts of biological consciousness and emerging questions about artificial consciousness, while maintaining scientific rigour.

Shevlin answers “no” to question 1, but admits a greater role for the general public in settling questions about AI consciousness.

Wiese argues that question 1 should be replaced with a more specific question, which will allow science to meaningfully contribute to questions about AI consciousness.

Elamrani defends “yes” in order to balance the arguments presented on question 1, while emphasizing the public’s key role in the societal integration of increasingly convincing AI as science slowly progresses towards settling the debate.

Talk Abstracts

Three Levels of Distinction Between Biological and Artificial Consciousness

Aïda Elamrani

Institut Jean Nicod, ENS-PSL

Consciousness is a complex phenomenon with manifest biological ties, which we analyse across three abstraction levels. At the physical level, a biological substrate, composed of cells and DNA, is often invoked as a plausible demarcation between conscious and non-conscious beings. At the phenomenal level, biological constraints such as embodiment, survival, or reproduction profoundly shape psychology and subjective experience. At the cognitive level, the range of input-output relations computed by conscious processes is bound by biological algorithms such as feedforward neural networks or Hebbian learning.

We argue that if the imitation or replication of any one of these levels is required for consciousness, then artificial systems remain far from achieving the kind of subjective states humans experience. As AI systems grow more sophisticated, distinct terminology may be necessary to differentiate advanced cognition from biological consciousness.

While the scientific question of consciousness is unlikely to be resolved within the next decade, increasingly advanced artificial agents will proliferate, raising pressing public questions about their integration and our tendency to attribute consciousness to them.

Digital Minds? Reconciling Folk Psychology and Consciousness Science

Henry Shevlin

University of Cambridge

Even as the cognitive sophistication of AI systems improves at a rapid pace, debates around AI consciousness remain mired in metaphysical and methodological controversies. From substrate neutrality to multiple realisation, the question of whether any artificial system could have subjective experience remains hotly contested, with little common ground between AI consciousness liberals and sceptics.

In this talk, I argue that we should not expect these disputes to achieve consensus resolution through new empirical evidence or novel philosophical arguments alone. Instead, I suggest that as the depth and complexity of human-AI interactions increases, intuitive priors among both experts and the general public are likely to increasingly favour theories willing to countenance artificial consciousness.

A further question is whether such shifts in sentiment can be rationally justified. I outline the case both for and against the role of intuitions in consciousness science, while emphasising that abandoning intuitions altogether would require a radical overhaul of current methods.

From Metaphysical Disputes to Open Empirical Questions: AI Consciousness as a Scientific Problem

Wanja Wiese

Ruhr-University Bochum

Can science meaningfully contribute to questions about AI consciousness? If yes, does this presuppose a judgment call about fundamental metaphysical questions regarding the nature of consciousness, for instance by committing to computational functionalism?

I suggest treating these questions themselves as empirical scientific questions. To answer them, we first have to identify relevant fundamental issues about the nature of consciousness, and then determine to what extent scientific research may resolve these issues. I submit that these issues include the following:

1. What features of living organisms are required for consciousness?
2. What differences between physical agents in non-virtual environments and virtual agents in computer simulations, if any, matter for consciousness?

I argue that science can meaningfully contribute to these questions by providing rigorous accounts of potentially relevant features of living organisms, and of differences between virtual and non-virtual agents. Specifying these features and differences, ideally mathematically, is the first step towards answering questions like 1 and 2. I illustrate this with examples from mathematical consciousness science.

Whether ultimate answers to such questions can be provided by science alone is currently still an open question. But science can at least contribute to answers by making these questions more intelligible and tractable.