

Understanding Artificial Consciousness: Clarifying Concepts and Advancing Research

Aïda Elamrani

ENS-PSL; IJN, DEC

Artificial Consciousness is a burgeoning field, promising significant advancements in modelling both artificial intelligence and human cognition. However, this interdisciplinary domain grapples with nascent and ambiguous terminology, often leading to misunderstandings and apprehensions. The first aim of this talk is to clarify the discourse by proposing key definitions and elucidating their historical and scientific context. The second aim is to enhance technical understanding of current research in the field, particularly focusing on implementations of the Attention Schema Theory and Transformer architectures. Building on this foundation, the last part of the talk seeks to alleviate fears surrounding AC and promote further research by highlighting the key advantages these models offer.