Conscious Machines: Facts, Fictions, and Fears

Robert Kozma

Center for Large-Scale Integrated Optimization and Networks
Department of Mathematics, University of Memphis, TN, USA

Abstract

For centuries, the question of consciousness has been considered to be so difficult and hard to address, that it was commonly assumed to belong to the realms of abstract philosophical speculations and metaphysics, and that it cannot be studied by rigorous empirical methods of scientific inquiries. The proliferation of brain monitoring technologies towards the end of the 20th century, however, produced compelling experimental results characterizing various aspects of conscious cognition and perception.

- New brain imaging results pointed to the feasibility of quantifying **neural correlates of consciousness (NCC)**. Along with the rapidly improving EEG, MEG, fMRI, PET, and other brain imaging techniques, numerous theoretical approaches have been emerged to interpret the experimental findings.
- **Global Workspace Theory (GWT)** is being one of the first prominent models, followed by Global Neural Workspace (GNW), Integrated Information Theory (IIT), Higher-Order Thought (HOT), Recurrent Processing Pathways (RPP), and others.
- Today, we do not only possess a rich reservoir of experimental and theoretical tools of human consciousness studies, but there are popular attempts to extend the studies to the realms of AI/AGI entities, possibly exhibiting **elements of Machine Consciousness**.

This talk provides an overview of the **state-of-art of experimental and theoretical studies** to human consciousness. We describe experimentally verifiable results on NCC and corresponding theories, as well as novel hypotheses which require further verification in the future. Several research groups used the NCC criteria derived originally for humans, to evaluate the behavior of AI/AGI agents. Some of the results show that certain criteria for NCC are satisfied in computer artifacts. The potential **emergence of conscious machines** raises anxiety and sometimes fears in the public. We provide a critical overview of the actual results and potential future developments. Specifically, we point out that today's AI designs lack many crucial aspects of intelligence, and **AI artifacts are far from matching human higher cognition and consciousness.**