

NOVEMBER 2017

AZRIELI PROGRAM IN BRAIN, MIND &  
CONSCIOUSNESS AT CIFAR

# IS TODAY'S ARTIFICIAL INTELLIGENCE ACTUALLY CONSCIOUS? NOT JUST YET.

*The computations that underpin current artificial intelligence (AI) more closely resemble unconscious processing than conscious thought in the human brain, suggesting that AI does not yet possess consciousness. Two particular processes that work synchronously in the human brain may provide key insights.*

## KEYWORDS

consciousness, self-monitoring, global availability, artificial intelligence, machine

## FOCUS OF REVIEW

To find out whether today's machines demonstrate consciousness, it's necessary to define consciousness in human and animal brains, then determine whether these qualities are present in AI. **Stanislas Dehaene** (Collège de France and Inserm-CEA) and **Sid Kouider** (École Normale Supérieure), both Senior Fellows in the **Azrieli program in Brain, Mind & Consciousness** at CIFAR, and **Hakwan Lau** (University of California, Los Angeles and University of Hong Kong) apply a computational lens to this question, reviewing research and providing insights into how human and animal brains process information and comparing the findings to the information processing methods used in today's AI. The authors seek to foster progress toward artificial consciousness, highlighting features that allow the human brain to generate consciousness so that these insights may be transferred into computer algorithms.

## BACKGROUND:

The prospect of machine consciousness has enthralled scientists and sci-fi enthusiasts for decades.

Recent advances in AI have driven progress toward this goal, with the creation of robots that can learn, solve complex problems and adapt to novel environments. However, the question persists of whether the subjective experience of consciousness escapes a computational definition — and whether machine architectures can produce such an experience in AI.

While advances in AI move the field forward, cognitive scientists have been decoding the enigma of consciousness in humans. Researchers have found that consciousness does not rely on a single mechanism, but many different types of computation. Moreover, research reports that the brain processes a vast proportion of information from the environment using unconscious mechanisms, and this information impacts the conscious experience.

The authors review and discuss the various dimensions of how the human brain computes information and how those relate to unconscious or conscious information processing, as follows:

- **Unconscious computation:** Neuroimaging research has revealed that most regions of the brain can be activated unconsciously. Meaning extraction, visual recognition of words or faces, reinforcement learning, cognitive control, and decision-making are a few examples of unconscious processing in humans. These strongly impact conscious processing.
- **Global availability computation** is the selection of external information to become available to the brain's many specialized subsystems for further processing. Research has associated global availability processing with a wide range of computations, including recollection, visual illusions, and serial information-processing. These systems converge toward making a single decision. Essentially, it makes possible coherent, thoughtful planning.
- **Self-monitoring computation** is the brain's ability to monitor its own knowledge and abilities, such as confidence in decisions, reflection, error detection, monitoring the quality of memory representations (meta-memory), and distinguishing reality from perception. This type of processing is also called introspection or meta-cognition.
- **Presence of consciousness:** Global availability and self-monitoring computations can operate separately, but it's only when the two work together that consciousness is present in humans. Case studies involving patients who have lost one of these abilities also describe a loss of conscious awareness, suggesting both are necessary for complete human consciousness.

These features of conscious thought challenge the idea that the subjective experience of consciousness escapes a computational definition.

## KEY INSIGHTS

Informed by research indicating that global availability and self-monitoring computations work together to produce consciousness in humans, the authors infer that consciousness is not yet present in AI.

- **Unconscious computation:** For the most part, information processing in today's AI involves computations that are similar to unconscious processing in the human brain. This includes AI functions such as speech-to-text, facial recognition, learning by reinforcement, and the current types of computations used for decision-making.
- **Global availability & self-monitoring computation:** Machine processing is advancing in global availability computations and self-monitoring computations, but there is still much progress to be made. A system's ability to undertake self-monitoring processes and to implement and flexibly coordinate multiple computations remains a challenge in machine architecture.
- **Presence of consciousness:** While progress is underway in both types of information processing, they are not yet joined to operate together. This achievement would produce a machine that behaves as though it were conscious.

## CONCLUSION AND IMPLICATIONS

To tackle machine consciousness, Dehaene, Lau, and Kouider argue that it's necessary to analyze the specific types of computations that make up consciousness in humans. They suggest that consciousness may not be fully explained by other current models.

The authors note that the empirical evidence reviewed is compatible with the possibility that specific computations are what produce consciousness in humans. Current AI platforms are built on sophisticated computations, but most of these are equivalent to unconscious computational processes in the human brain and not to those believed to produce consciousness.

To zero in on machine consciousness, researchers should aim to generate computational processes similar to those that give rise to two main types of consciousness in humans: global availability consciousness and self-monitoring. Developing similar types of information processing mechanisms in AI may lay the foundation for consciousness in machines.

## REFERENCES

Stanislas Dehaene et al., What is consciousness, and could machines have it? *Science* 2017;358:issue 6362.

## RESEARCH BRIEF AUTHORSHIP

Stephanie Orford, BSc (Honours) Behavioural Neuroscience, Simon Fraser University.

## AZRIELI PROGRAM IN BRAIN, MIND & CONSCIOUSNESS AT CIFAR

This pioneering program examines the neural underpinnings of consciousness, bridging disciplines as diverse as cognitive science, computer science and philosophy to understand consciousness. Insights produced through this interdisciplinary global research program can help us understand subjective experience, create better pharmaceutical treatments, and explore new ways of interacting with machines.

# CIFAR

CANADIAN INSTITUTE FOR ADVANCED RESEARCH  
MaRS Centre, West Tower  
661 University Ave., Suite 505  
Toronto, ON M5G 1M1

[www.cifar.ca](http://www.cifar.ca)