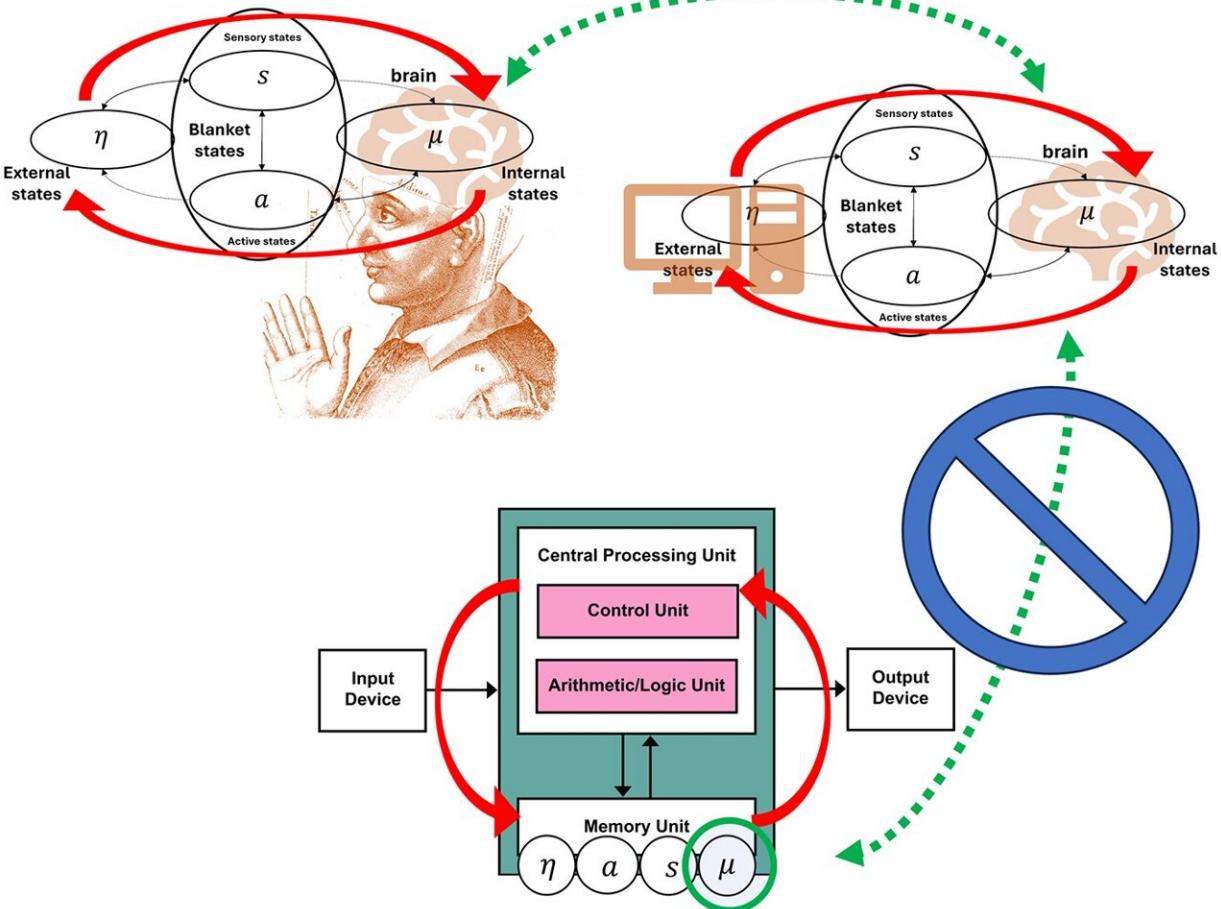# Can consciousness exist in a computer simulation?

July 19 2024, by Meike Drießen

A paradigmatically embodied individual (a brain within a body; top left), as well as a brain in a vat (top right), which can be regarded as a limiting case of a paradigmatically embodied individual. A part of a computer simulation that simulates a conscious being (bottom), however, is not a limiting case of such an individual, because it could not become incorporated into a paradigmatically

Would it be desirable for artificial intelligence to develop consciousness? Not really, for a variety of reasons, according to Dr. Wanja Wiese from the Institute of Philosophy II at Ruhr University Bochum, Germany.

In an essay, he examines the conditions that must be met for consciousness to exist and compares brains with computers. He has identified significant differences between humans and machines, most notably in the organization of brain areas as well as memory and computing units.

"The causal structure might be a difference that's relevant to consciousness," he argues. The essay was published on June 26, 2024 in the journal "*Philosophical Studies.*"

Two approaches

When considering the possibility of consciousness in artificial systems, there are at least two different approaches.

One approach asks: How likely is it that current AI systems are conscious—and what needs to be added to existing systems to make it more likely that they are capable of consciousness? Another approach asks: What types of AI systems are unlikely to be conscious, and how can we rule out the possibility of certain types of systems becoming conscious?

In his research, Wanja Wiese pursues the second approach.

"My aim is to contribute to two goals: Firstly, to reduce the risk of inadvertently creating artificial consciousness; this is a desirable outcome, as it's currently unclear under what conditions the creation of artificial consciousness is morally permissible. Secondly, this approach should help rule out deception by ostensibly conscious AI systems that only appear to be conscious," he explains.

This is particularly important because there are already indications that many people who often interact with chatbots attribute consciousness to these systems. At the same time, the consensus among experts is that current AI systems are not conscious.

The free energy principle

Wiese asks in his essay: How can we find out whether essential conditions for consciousness exist that are not fulfilled by conventional computers, for example? A common characteristic shared by all conscious animals is that they are alive.

However, being alive is such a strict requirement that many don't consider it a plausible candidate for a necessary condition for consciousness. But perhaps some conditions that are necessary for being alive are also necessary for consciousness?

In his article, Wanja Wiese refers to British neuroscientist Karl Friston's free energy principle. The principle indicates: The processes that ensure the continued existence of a self-organizing system such as a living organism can be described as a type of information processing.

In humans, these include processes that regulate vital parameters such as body temperature, the oxygen content in the blood and blood sugar. The same type of information processing could also be realized in a computer. However, the computer would not regulate its temperature or

blood sugar levels, but would merely simulate these processes.

Most differences are not relevant to consciousness

The researcher suggests that the same could be true of consciousness. Assuming that consciousness contributes to the survival of a conscious organism, then, according to the free energy principle, the physiological processes that contribute to the maintenance of the organism must retain a trace that conscious experience leaves behind and that can be described as an information-processing process.

This can be called the "computational correlate of consciousness." This too can be realized in a computer. However, it's possible that additional conditions must be fulfilled in a computer in order for the computer to not only simulate but also replicate conscious experience.

In his article, Wanja Wiese therefore analyzes differences between the way in which conscious creatures realize the computational correlate of consciousness and the way in which a computer would realize it in a simulation. He argues that most of these differences are not relevant to consciousness. For example, unlike an electronic computer, our brain is very energy efficient. But it's implausible that this is a requirement for consciousness.

Another difference, however, lies in the causal structure of computers and brains: In a conventional computer, data must always first be loaded from memory, then processed in the central processing unit, and finally stored in memory again. There is no such separation in the brain, which means that the causal connectivity of different areas of the brain takes on a different form. Wanja Wiese argues that this could be a difference between brains and conventional computers that is relevant to consciousness.

"As I see it, the perspective offered by the free energy principle is particularly interesting, because it allows us to describe characteristics of conscious living beings in such a way that they can be realized in artificial systems in principle, but aren't present in large classes of artificial systems (such as computer simulations)," explains Wanja Wiese.

"This means that the prerequisites for consciousness in artificial systems can be captured in a more detailed and precise way."

More information: Wanja Wiese, Artificial consciousness: a perspective from the free energy principle, *Philosophical Studies* (2024). [DOI: 10.1007/s11098-024-02182-y](DOI: 10.1007/s11098-024-02182-y)

Provided by Ruhr-Universitaet-Bochum