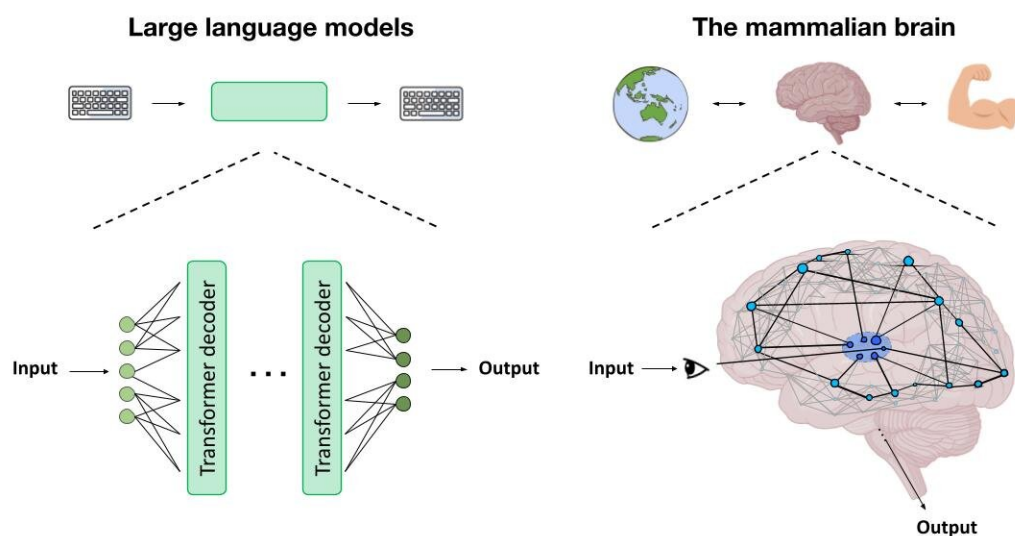# Applying a neuroscientific lens to the feasibility of artificial consciousness

October 27 2023

Left: a schematic depicting the basic architecture of a large language model, which can havetens or even more than a hundred decoder blocks arranged in a feed-forward fashion.Right: a heuristic map of the thalamocortical system, which generates complex activity patterns thought to underlie consciousness. Credit: Mac Shine, Jaan Aru

The rise in capabilities of artificial intelligence (AI) systems has led to the view that these systems might soon be conscious. However, we might

be underestimating the neurobiological mechanisms underlying human consciousness.

Modern AI systems are capable of many amazing behaviors. For instance, when one uses systems like ChatGPT, the responses are (sometimes) quite human-like and intelligent. When we, humans, are interacting with ChatGPT, we consciously perceive the text the [language](#) model generates, just as you are currently consciously perceiving this text here.

The question is whether the language model also perceives our text when we prompt it. Or is it just a zombie, working based on clever pattern-matching algorithms? Based on the text it generates, it is easy to be swayed that the system might be conscious.

New research [published](#) in the journal *Trends in Neurosciences*, Jaan Aru, Matthew Larkum and Mac Shine take a neuroscientific angle to answer this question.

The three neuroscientists argue that although the responses of systems like ChatGPT seem conscious, they are most likely not. First, the inputs to language models lack the embodied, embedded information content characteristic of our sensory contact with the world around us. Secondly, the architectures of present-day AI algorithms are missing key features of the thalamocortical system that have been linked to [conscious awareness](#) in mammals.

Finally, the evolutionary and developmental trajectories that led to the emergence of living conscious organisms arguably have no parallels in artificial systems as envisioned today. The existence of living organisms depends on their actions and their survival is intricately linked to multi-level cellular, inter-cellular, and organismal processes culminating in agency and consciousness.

Thus, while it is tempting to assume that ChatGPT and similar systems might be conscious, this would severely underestimate the complexity of the neural mechanisms that generate consciousness in our brains. Researchers do not have a consensus on how consciousness rises in our brains. What we know, and what this new paper points out, is that the mechanisms are likely way more complex than the mechanisms underlying current language models.

For instance, as pointed out in this work, real neurons are not akin to neurons in artificial neural networks. Biological neurons are real physical entities, which can grow and change shape, whereas neurons in large language models are just meaningless pieces of code. We still have a long way to go before we understand consciousness and, hence, a long way to conscious machines.