Review of: "The functional unit of neural circuits and its relations to eventual sentience of artificial intelligence systems"

Scythia Marrow

Potential competing interests: No potential competing interests to declare.

The article has a great central thesis, one which I agree with, that consciousness requires learning over complex dynamics and cannot be replicated using static methods such as gradient descent over differentiable functions (such as deep neural networks).

The article does not support the thesis using any acceptable scientific evidence, makes several claims that are patently false, and fundamentally misrepresents both current machine learning approaches as well as its own contributions. It has no mathematical grounding, and the single mathematical idea it puts forward, that of biases in threshold logic units (which I will be referring to as activation functions, as per the machine learning literature) being a natural way to create recurrence relations by assigning it as a set-point signal instead of a learned weight, is simply cited as is with no further mathematical elaboration. So it is not novel. This work is also mathematically equivalent to a long short-term memory unit, which has been a key component of DNNs for several years[1][2]. So the author's assertion that “Since the discussion on the possibility of achieving self-awareness by currently constructed artificial intelligence systems is necessary and will be continued, drawing attention here to the properties of so-called functional units of neural circuits is justified” is incorrect, because their main contribution to the properties of neural circuits has already been identified by the machine learning literature. Indeed, the machine learning literature has moved far beyond the author's issues with recurrence: graph neural networks and boltzmann machines both are fully recurrent, graph neural networks rely on gradient descent over a stationary fixed point[4], but boltzmann machines can learn without gradient descent and over any energy function[5]. Finding a mathematical reason for why boltzmann machines with an arbitrary but static energy function cannot learn self-visualization is a challenging mathematical problem and not a given as the authors may assume.

So the authors fundamentally misunderstand current machine learning research, and have no contribution to make to the field. Yet they also fail to account for common knowledge within cognitive psychology. The authors claim that “it appears that the crux of self-awareness lies in the capacity to generate a self-image” which is incorrect, as recent work on aphantasia demonstrates that people can be conscious, aware, and living normal, healthy lives without the ability to generate a self-image [3]. If the authors wish to claim that they mean self-image in a more metaphorical way rather than literal mental imagery they should both move away from the terminology they are using and also not cite so many studies on mental imagery.

The authors also claim that the electromagnetic field (EMF) created by the brain is integral to the creation of consciousness. Somehow. They do not explain why or how, although they cite several sources concerning the quantum brain hypothesis, which has actually has some promising recent results that hold up to some scrutiny[6][7][8]. But the quantum brain hypothesis is an extraordinary claim, and requires strong evidence that the authors do not provide. Their contribution to the hypothesis, that local recurrence relations create a global EMF which entangles disparate parts of the brain by interacting with micro-tubules, breaks physics. Entanglement is a local phenomenon and a global EMF cannot be the cause of the quantum brain. There is no evidence that a global EMF is required for consciousness, and claiming it does breaks either fundamental physics or the Church-Turing thesis.

I do not recommend this article for publication without serious reworks. The main thesis needs to be supported by mathematical reasoning and computational evidence that takes into account and builds on the standards of modern machine learning theory. The background sections on both machine learning literature and cognitive psychology literature need serious expanding and elaboration. The authors need to remove false claims about mental imagery and either strongly elaborate on how their proposed EMF solution to consciousness doesn't contradict known physics or remove that as well.

Finally, the structure needs work. The article is meandering and uses too many pages to say very little. It doesn't talk about the main thesis before the very last paragraph, uses meandering “woo” language that is not scientifically rigorous, appeals to intuition instead of mathematical grounding, and has complex figures with which do not seek to explain the main thesis.

[1] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.

[2] Medsker, L. R., & Jain, L. C. (2001). Recurrent neural networks.*Design and Applications*, *5*(64-67), 2.

[3] Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia.*Cortex*, *105*, 53-60.

[4] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model.*IEEE transactions on neural networks*, *20*(1), 61-80.

[5] Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines.*Cognitive science*, *9*(1), 147-169.

[6] Kerskens, C. M., & Pérez, D. L. (2022). Experimental indications of non-classical brain functions.*Journal of Physics Communications*, *6*(10), 105001.

[7] Warren, W. S. (2023). Comment on:‘Experimental indications of non-classical brain function’2022 Journal of Physics Communications 6 105001. *Journal of Physics Communications*, *7*(3), 038001.

[8] Kerskens, C., & Pérez, D. L. (2023). Reply to Comment on:‘Experimental indications of non-classical brain function’(2022 Journal of Physics Communications 6 105001). *Journal of Physics Communications*, *7*(3), 038002.