

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

Kieran Greer

Potential competing interests: No potential competing interests to declare.

I think that this is a very interesting paper. It is certainly based on the real brain and is an interesting neural system at least. Figure 3a is very remarkable. Does it need to re-learn each time the set point is changed? I would suggest however that backpropagation for error correction is different to a repeating feedback loop. The error correction, for example, happens only once each iteration. OK, so it can repeat until it stabilises to some value, but that is a self-organising neural network. I do not think that the system is complete yet. One question is how it oscillates between the two different regions, when the feedback is local only. If it is through local interactions, then the signal sign needs to change, as you indicated in the theory, or if it is between the two regions, then some type of matching process is required so that they can achieve the same signal values.

As you state, backpropagation knows the output and so it can produce an accurate error value. Your feedback loop does not know this. If it is correcting to match with a pattern, then how does the system know what the correct pattern to match with is? If it requires an exhaustive type of search to find a good match, then how would that be done. Maybe at the neuronal level, the number of exhaustive combinations can be relatively small. You then make the point just below figure 3, when describing the threshold logic unit.

'Comparison of the outline of the feedback loop and another, old known notion of the threshold logic unit reveal that they are related to each other.'

So if there is already an association between the two neural layers, they will be activated by that relation. Then as you state, a better match will tend the value to 1, or its maximum value. But there are still questions about what exactly has been learned. How many of these units are required, for example and are they also the input patterns?

But the system oscillations are to highlight part of the system and so start after a better match. Error correction would use oscillations to stabilise and then switch them off. So ultimately, it will oscillate between, say, plus or minus 1 or some threshold, when it may start with smaller values like plus or minus 0.5?

What is the similar EMF field that exists in AI systems. They are written in software, so it must be a programmed feature. I think that the emergence of a conscious is dubious with the current hardware of the AI systems, but the emergence of some type of AGI might be possible. ChatGPT, for example, with billions of nodes, does things that cannot be understood. That may have emerged through the large number of neurons involved. But you write about self-awareness, so is that a



sense or an intelligence idea?