

The only game in town

In a recent workshop at the Conference on Neural Information Processing Systems (NeurIPS), future directions at the intersection of neuroscience and AI were considered. A panel discussion at the end of the day started with a provocative question: do we need AI to understand the brain?

It may seem obvious that artificial intelligence (AI) would look to neuroscience for inspiration. After all, if one wants to design a machine that performs a feat of intelligence — such as an algorithm to categorize visual images — why not base it on the visual system of the brain?

Indeed, since the mid-twentieth century, neuroscience has inspired AI¹. For example, in 1943 McCulloch and Pitts proposed ‘nervous nets’ consisting of artificial neurons². In 1948, Turing wrote a prescient, but unpublished in his lifetime, paper on ‘Intelligent machinery’ with a section on ‘The cortex as an unorganized machine’³. The properties of neurons in the visual system⁴ inspired the hierarchical representations of deep learning architectures.

But what is biological inspiration? The brain is a jumble of complex biological stuff, the vast majority of which we don’t understand. So, how does one go from biology to AI? Typically, an AI researcher attempts to discern the principles by which the brain implements a specific function, in order to then try to implement it in a machine. That is, an AI researcher seeks the principles used by the brain at the algorithmic level, such as the general way in which different cell types in the visual cortex process simple stimulus features to create more complex representations.

Traditionally, neuroscientists and AI researchers have had different goals and different ways of understanding, even if the phenomena or tasks of interest are similar, such as vision. Neuroscientists try to explain, in mechanistic and computational detail, how various processes are implemented in the brain. Face perception, for example, is a fundamental ability that most humans perform well, and usually automatically. Although much is known about the neuroscience of face perception, a great deal is not known, and hence it is an area of active research, one that will probably be investigated for decades to come. It may not

be possible, then, for neuroscientists to get a full mechanistic picture of a function like face perception, at least in the short term. At a recent panel discussion at NeurIPS in Vancouver on ‘A new hope for neuroscience’, it was pointed out that what goes on at the single-cell level may be almost inscrutable. Therefore, in addition to getting more data, another approach is to turn the tables and look to AI for inspiration to better understand the complexity of the brain.

The panel discussion was lively and provocative. The seven panellists were Cristina Savin (NYU), Tim Lillicrap (Google DeepMind), David Sussillo (Google Brain), Yoshua Bengio (Université de Montréal/Mila), Blake Richards (McGill/Mila), Surya Ganguli (Stanford), and Ila Fiete (MIT). The moderator (Guillaume Lajoie, Université de Montréal) started by asking whether the following statement is true or false: “To truly understand the brain, one needs to study artificial intelligence.”

Four panellists responded in the affirmative, but the others were less sure. Surya Ganguli, who was cautious, commented first: “What current AI systems are generating are the only game in town for computational models that we’ve identified that actually solve problems.” However, he warns “we shouldn’t overfit our intuition on those, and should be open-minded about other possibilities.” Blake Richards, an affirmative responder to the question, said, “If the goal of artificial intelligence is to engineer intelligent systems, and the goal of neuroscience is to understand the biological mechanisms that underpin animal intelligence, then surely we need each other in order to get at those general principles that apply to both.” David Sussillo commented: “We could be misled by these artificial systems and we need to be careful about that.” The conversation continued, and the topic seems ripe for exchanging ideas.

A further question for the panellists was to comment on two categories of research: studying the computations in brain networks

(and possibly in artificial networks), versus studying the learning rules that shape those computations. Yoshua Bengio argued that the whole field of machine learning tells us that particular kinds of computations do their job in the context of learning. Ila Fiete replied that learning rules are important but can’t be the whole story, because there is a lot of structure and modularity in the brain, including inductive biases. Surya Ganguli added that the brain may have solved the learning problem with specialized architectures and simple local learning rules that only work for those neural architectures, and that might be understandable in a way that current artificial systems are not. Tim Lillicrap countered that evolution probably favoured robust learning rules for solving complex problems. The interested reader can watch the full panel discussion [here](#) and read more about the workshop where the panel took place [here](#).

Clearly, there is enthusiasm for exploring synergies between AI and neuroscience — a welcome development. Recent review articles^{5,6} and interdisciplinary conferences (for example, [Cognitive Computational Neuroscience](#)) attest to this enthusiasm. But challenges remain. Progress in neuroscience is slow; progress in deep learning is fast, at least in the past decade within certain domains. Interdisciplinary conversations and collaborations are essential, as exemplified in this workshop. Let the dialogue continue. □

Published online: 14 February 2020
<https://doi.org/10.1038/s42256-020-0158-5>

References

- Schmidhuber, J. *Neural Netw.* **61**, 85–117 (2015).
- McCulloch, W. S. & Pitts, W. *Bull. Math. Biophys.* **5**, 115–133 (1943).
- Turing, A. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus the Secrets of Enigma* (ed. Copeland, J.) Ch. 10 (Clarendon, 2004).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106–154 (1962).
- Kriegeskorte, N. & Douglas, P. K. *Nat. Neurosci.* **21**, 1148–1160 (2018).
- Richards, B. A. et al. *Nat. Neurosci.* **22**, 1761–1770 (2019).