Artificial and Natural Intelligence: From Invention to Discovery

Li Zhaoping[1,2,*]
[1]University of Tübingen, Tübingen, Germany
[2]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
*Correspondence: li.zhaoping@tuebingen.mpg.de
https://doi.org/10.1016/j.neuron.2020.01.014

An international group of researchers met in November 2019 in Beijing to explore the intersection of neuroscience and AI. The aim was to offer a fertile ground for stimulating discussions and ideas, including issues such as policy making and the future of neuroscience and AI across the globe.

What is a better model of the human brain: a mouse brain or an engineered artificial intelligence (AI) system? This was one of many questions discussed by brain scientists and AI engineers during the thought-provoking November 2019 "AI and the Brain" conference in Beijing, directed by Cell Press editors Mariela Zirlinger and Moran Furman and organized as part of a partnership between Cell Press and the Beijing Municipal Science and Technology Commission.

With more brain scientists than AI engineers being present at the meeting, many of the discussions naturally centered on how brain science could benefit from AI and how state-of-the-art AI systems can be interpreted in the context of current understandings in brain science. This has at least two facets. One is uncontroversial: the observation that methods of machine learning (ML), which is almost synonymous with modern AI, are advancing progress in all natural sciences, including brain science. The second, which we consider first, is more controversial and asks whether there is a special connection between AI and brain science.

Machines have, for decades, been able to do better than humans in many tasks that would once have been considered to require some form of intelligence, such as multiplying two multi-digit numbers. More recently, they have bested us in ever more challenging cases such as playing chess, Go, or StarCraft. Thus, it is just a question of when, and in which ways, AI will supplant additional facets of human intelligence, an issue perhaps of more interest to the general public and science fiction authors than to the meeting participants. The attendees were, of course, able to point to many areas for improvement in current technologies, particularly in tasks humans find easy, such as washing dishes, playing soccer, learning concepts or tasks from a few examples, and recognizing partially occluded objects in cluttered scenes or out of context (imagine a penguin in a rain forest). Current AI often lacks robustness; it can be easily fooled so that, for example, it recognizes a panda in an image as a gibbon when the image pixel values are modified so slightly that the modification is almost imperceptible to humans. Furthermore, current AI systems typically consume vastly more power than a human brain for performing the same task.

Meanwhile, the recent rapid progress of AI has prompted some to wonder whether AI could move forward at a fast pace with little, or at least limited, input from brain science. What would progress look like in AI without brain science? Is there anything uniquely interesting about natural, i.e., biological, intelligence? It is important to remember that AI and brain science have different cultures and value systems. In particular, brain science (using experimental methods from, e.g., physiology, anatomy, psychology, and medicine and theoretical/computational methods from, e.g., physics, computer science, and mathematics) is a discipline in *natural science* whose practitioners aim to *discover* nature's laws to understand the world, whereas AI is a discipline in *engineering* in which researchers *invent* and innovate technologies, often by applying nature's laws or guided by intuition and experience. Historically, corresponding science and engineering disciplines (e.g., chemistry and chemical engineering) typically helped to spur each other's progress. This should also happen between brain science and AI; one could even envision the formation of a joint discipline of intelligence, be it artificial or natural. Many researchers are active in both brain science and AI, while many have their preference for brain science or AI driven by their interest to discover or invent, respectively.

Because many of the participants work in artificial or biological vision, and because vision is one of the areas in which AI and brain science have had many interactions, it is most straightforward to consider the case of visual intelligence. In the primate brain, visual information is transformed from pixel inputs as activation of photoreceptors in the retina to neural responses in subsequent visual cortical areas, such as V1, V2, V3, V4, and IT, along the visual pathway. However, even though we are able to record neural activities at multiple stages of this pathway, we are surprisingly ignorant about the exact form of the transformations that lead to our generally exemplary performance at tasks such as object recognition. This ignorance is particularly severe in later stages of the pathway.

Mimicking the hierarchical structure of the visual pathway, the most successful AI vision networks are built as convolutional neural networks of many layers. The values of the (often) millions of parameters defining the precise transformations are determined by a form of so-called supervised learning involving an instance of the chain rule of calculus known as backpropagation. The idea is to present a large number of visual images to the input layer of the network and to teach the network such that its upper layer can correctly signal or identify the objects in the images

(Yamins and DiCarlo, 2016). If one excludes cases when the AI vision can be fooled or be confused, some AI networks can already achieve human or even super-human performance in object recognition.

If we can invent an intelligent technology in AI, we can ask whether we can find the same or corresponding computational principle or algorithm in a biological brain, allowing that the actual substrates that implement the algorithms, whether in silicon or in "wet" neurons, could differ. Could we conjecture that if the AI algorithm is sufficiently ingenious (while using limited resources), then there must be a biological counterpart? If so, the AI vision network, for example, could guide us in understanding the visual signal transformations along the visual pathway (and conversely, AI practitioners could "invent" better solutions by borrowing brain's algorithms). James DiCarlo, a vision scientist from MIT, noted that indeed, among the AI networks that do not employ too many computational layers, the ones that perform better at object recognition have their computational units respond to visual inputs in the same or in a similar way as neurons in monkey visual cortex, so that the biological neural responses can be related in a simple linear way to the responses in the artificial neurons (Yamins and DiCarlo, 2016). Mu-ming Poo, director of the Institute of Neuroscience in Shanghai, found a form of plasticity in neural synapses resembling backpropagation. Since we can examine the artificial networks more easily to see the pathways and transformations from the visual input sensors to individual nodes in various layers of the networks, the AI networks can hopefully help us to hypothesize or understand how our brain builds a representation of the visual world layer by layer, from the original raw image pixels on retina to neurons signaling complex objects, such as faces, in higher visual cortex.

However, compared to biological vision, current AI visual networks have irksome problems: the ease of fooling them and their fragility to image context, scene clutter, and even partial occlusion. Such shortcomings are not easily overcome, for instance, by mimicking the biological neural spikes in artificial neurons, reported Jun Zhu from Tsinghua University. Somehow, current AI vision lacks what one might call visual "understanding."

Many scientists (including me) suggest that this "understanding" is represented in the top-down feedback from higher to lower visual cortical areas along the visual pathway in the brain (Zhaoping, 2019). Such feedback, absent in the best AI vision networks so far, realizes this "understanding" by being able to synthesize the would-be visual signals in the lower visual areas for various objects. Synthesis of this sort constitutes internal knowledge that can help to fill in missing visual input signals due to occlusion and to dismiss misleading input signals that are inconsistent with the top-down knowledge.

Wu Li from Beijing Normal University shared lessons his team learned from feedback from V4 or V2 to V1 in monkey brains (Chen et al., 2014). DiCarlo, Gabriel Kreiman (from Harvard), and their collaborators have seen a signature of such feedback in the higher visual cortical areas of primates: longer latencies in neural responses to more challenging visual inputs, potentially due to the interaction between top-down feedback and bottom-up feedforward signals.

Another important facet of human vision that is incompletely reflected in AI systems is the critical role played by attentional selection so that only a fraction of sensory inputs is further processed by our brain (not the least saving power and processing resources). Theoretical arguments, backed up by behavioral data, turn this facet into the prediction that top-down feedback is likely directed mainly to the central visual field. Indeed, human vision in peripheral visual fields (which, according to this hypothesis, would lack the benefit of top-down synthesis) is, like AI, also easily fooled, by being more susceptible to visual illusions (Zhaoping, 2019).

Extending beyond vision, meeting attendees pondered over many of the brain's peculiarities that are novel in AI. Poo and others emphasized that the brain has lateral connections within individual brain areas, in addition to the feedforward and feedback connections, and Virginia de Sa of the University of California San Diego shared her insights from using all these connections in her neural network models. Poo also emphasized the diversity in neuronal types and that it is easier to strengthen or weaken an existing neural connection through learning than to create a new connection between neurons (Poo et al., 2016). Shimon Ullman from the Weizmann Institute argued that evolution may have built in some critical elements of the neural network architecture that enables babies' brains to learn in a way that the current AI could not. György Buzsáki of New York University went even further, suggesting, based on his studies, that the brain, unlike AI, has rich pre-configured neural dynamics and that the outside world can exert its influence for sensory responses and learning only through these dynamics. Minmin Luo of the Chinese Institute for Brain Research shared the wonders of neuromodulators such as dopamine and serotonin for reward learning. Daeyeol Lee of Johns Hopkins University and Xiaojing Wang of New York University marveled at the different timescales of neural dynamics across different brain regions and wondered whether they may be key to adapting to the environment at multiple timescales.

The physicist Richard Feynman famously said, "what I cannot create, I do not understand." Taking this literally, engineering robots to produce or mimic biological intelligence behaviorally, in a robust and energy-efficient manner, should help scientists identify critical and relevant questions that could otherwise be easily missed (such as when one looks at visual object recognition without looking at motor behavior for, or guided by, visual sensing). Along this line, Tony Prescott of Sheffield University presented work on robots with artificial, rodent-inspired whiskers. Yulia Sandamirskaya of ETH Zurich described neuromorphic cognitive robots that can, for example, manipulate a Rubik's cube. Sen Song of Tsinghua University introduced a neuromorphic chip architecture that enables researchers to try real-world intelligent tasks such as an unmanned bicycle riding on a road (Pei et al., 2019). Julie Grollier of CNRS/Thales France builds nanodevices to improve neuromorphic computing and energy efficiency, taking into account that, contrary to commonplace computing devices, memory and processing need to be together in the brain. Such endeavors highlight issues that need to be tackled: for example, spatial coordinate transformations between sensory and motor systems, autonomous learning, interactions between sensory processing, memory, and action control.

Let us turn back to the first facet of the link between AI and brain science mentioned at the opening: applying AI tools to the advancement of neuroscience research. Historically, invented tools, such as the telescope and microscope, have greatly accelerated the pace of scientific discovery. ML methods, whose development has been turbocharged by the burgeoning of modern AI, are tools that can greatly aid progress in brain science. In particular, the ability of these methods to process large datasets should help brain scientists to analyze and extract insightful information from the massive and multi-dimensional data that are becoming more and more common in brain science. Juan Zhou of National University of Singapore shared the example of her team using ML methods to fuse multimodal neuroimaging data in their study of neuropsychiatric disorders.

Indeed, the meeting also celebrated the success and power of ML to decode cognitive or behavioral signals from neural signals. For example, Edward Chang and colleagues at University of California, San Francisco, decode articulatory movements for producing speech from human sensorimotor cortical recordings, and Kafui Dzirasa and colleagues from Duke University use neural signals from multiple brain regions to decode human emotions. Such developments can not only create valuable medical applications, but also enable discoveries as to how the brain represents and processes information. For example, they could reveal whether speech production involves a brain part different from that used for language processing. This revelation can impact investigations such as those by another meeting attendee, Liping Wang of Institute of Neuroscience in Shanghai, on how human and monkeys process and represent language syntax.

To use ML methods well, researchers need to be properly trained, and this training is not as simple as the training needed to use a microscope. This calls for reforms in higher education. Some would say that it is becoming imperative for students in life science disciplines to acquire the sort of mathematics and computer science skills that are currently expected from physics and engineering students.

How to replicate the admirable pace of the development of modern AI in brain science also prompted much discussion. The importance of asking the right questions, formulating them properly, and communicating the ideas and findings clearly was noted. AI has certainly benefited hugely from concrete benchmarks, in particular public datasets on which performance of different AI solutions can be rigorously quantified and compared. Benchmarking, no doubt, has its drawbacks, as nicely reflected in Goodhart's law: "when a measure becomes a target, it ceases to be a good measure." Still, the very process by which this becomes true has repeatedly led to substantial and beneficial innovation in engineering and in AI in particular.

In engineering disciplines, the value of an invention, such as an AI solution, is its utility or performance (such as in benchmarks). In science, the value of a theory is only available and authentic through the agreement between the precise theoretical predictions and experimental data, and this value then resides in the theoretical insights and understanding that enable the predictions that otherwise cannot be easily extrapolated from pre-existing knowledge. The discussion thus turned to more general concerns. Brain scientists admonished themselves to adhere to the scientific value, so as to expect scientific theses to be clearly presented, in particular to provide precise and experimentally falsifiable theoretical predictions. In particular, the community can benefit from self-criticism in the style of the 20th century physicist Wolfgang Pauli, who reserved his most severe disapproval for ideas and claims that were "not even wrong" (Peierls, 1960) because they were so obscurely presented that they could not be evaluated or tested and thus did not belong within the realm of science. Both education and leadership should greatly help to nurture such a community culture in brain science.

ACKNOWLEDGMENT

I thank Peter Dayan and meeting attendees, particularly James DiCarlo, Wu Li, Tony Prescott, and Shimon Ullman, for their very helpful comments on the draft versions of this article and University of Tübingen and the Max Planck Society for funding.

REFERENCES

Chen, M., Yan, Y., Gong, X., Gilbert, C.D., Liang, H., and Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. Neuron *82*, 682–694.

Pei, J., Deng, L., Song, S., Zhao, M., Zhang, Y., Wu, S., Wang, G., Zou, Z., Wu, Z., He, W., et al. (2019). Towards artificial general intelligence with hybrid Tianjic chip architecture. Nature *572*, 106–111.

Peierls, R.E. (1960). Wolfgang Ernst Pauli, 1900-1958. Biogr. Mem. Fellows R. Soc. *5*, 174–192, https://doi.org/10.1098/rsbm.1960.0014.

Poo, M.M., Du, J.L., Ip, N.Y., Xiong, Z.Q., Xu, B., and Tan, T. (2016). China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing. Neuron *92*, 591–596, https://doi.org/10.1016/j.neuron.2016.10.050.

Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. *19*, 356–365.

Zhaoping, L. (2019). A new framework for understanding vision from the perspective of the primary visual cortex. Curr. Opin. Neurobiol. *58*, 1–10.