CellPress

Cell
Leading Edge

Commentary

The application of artificial intelligence to biology and neuroscience

Blake Richards,[1,2,3] Doris Tsao,[4,5] and Anthony Zador[6,*]
[1]McGill University, Montreal, QC, Canada
[2]Mila, Montreal, QC, Canada
[3]CIFAR, Toronto, ON, Canada
[4]University of California, Berkeley, Berkeley, CA, USA
[5]Howard Hughes Medical Institute, USA
[6]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
*Correspondence: zador@cshl.edu
https://doi.org/10.1016/j.cell.2022.06.047

Over the last decade, the artificial intelligence (AI) has undergone a revolution that is poised to transform the economy, society, and science. The pace of progress is staggering, and problems that seemed intractable just a few years ago have now been solved. The intersection between neuroscience and AI is particularly exciting.

What is AI?

The term artificial intelligence (AI) has no strict definition. Broadly speaking, AI refers to computer systems that are designed to mimic human intelligence, with the goal of performing any task that a human can perform (Figure 1). AI is generally considered a subfield of computer science but is closely allied with several other areas of research, including data science and machine learning, as well as statistics. Much of the promise of AI in the sciences derives from its ability to discover (or "learn") structure in large datasets and to use this structure to make predictions or even perform tasks. Such AI systems have strengths that can complement those of humans. For example, AI systems have the ability to see patterns in very high-dimensional data and thus can serve as powerful tools to assist rather than replace human researchers. Almost all modern AI systems rely on variations of artificial neural networks (ANNs), which were inspired by the organization of the nervous system.

There are three classic paradigms in AI for extracting structure from data. In "supervised learning," the data consist of pairs—an input item (e.g., an image of an elephant) and its label (e.g., the word "elephant")—and the goal is to predict the labels of novel items. Supervised learning can be seen as a particularly powerful form of nonlinear regression. In "unsupervised learning," the data have no labels, and the goal is to find the underlying statistical structure (e.g., to infer the existence of elephants and giraffes from a collection of safari pictures). Unsupervised learning can be seen as a generalization of classical statistical techniques such as clustering and principal component analysis. (Many modern AI systems also rely on "self-supervised learning," which achieves the same goals as unsupervised learning by labeling the data using automatic methods, e.g., applying the same label to different artificially generated variations of an object). Finally, in "reinforcement learning," the task is to discover strategies that achieve some goal, using information about the rewards obtained from previous actions. Reinforcement learning approaches have recently been used to achieve superhuman performance in games such as chess and Go, as well as in the design of novel drugs.

In what follows, we first discuss the impact of AI tools in analyzing and interpreting data on the life sciences, with special focus on neuroscience. We then focus on a second application of AI, specific to neuroscience, whereby ANNs are used as models for how biological neural networks compute. This commentary complements several other recent reviews on applications of AI to biology and medicine (Rajpurkar et al., 2022; Hassabis et al., 2017; Sapoval et al., 2022; Kriegeskorte and Douglas 2018).

AI tools for analyzing and interpreting data

The first important application of AI involves the development of tools for analyzing and interpreting data. For example, the motion tracking software Deeplabcut (Mathis et al., 2018), which relies on AI tools, now makes it possible to analyze video in order to identify and/or label the precise pose of animals, enabling much more precise characterization of animal behavior (both of individuals and social groups) during neural recordings or perturbations. Another application of machine learning is to reconstruct synaptic connectivity maps from serial electron microscopy data using segmentation and tracking algorithms from machine learning, which has so far resulted in reconstruction of an entire *Drosophila* brain, the entire mouse retina, and a cubic millimeter of mouse V1 (MICrONS Consortium, 2021). Such systems are transforming how experimental data are collected and interpreted.

Although the primary focus of this commentary is on the application of AI to neuroscience, AI has applications to many other areas of biology. For example, AI has important applications to diverse fields including protein modeling, the analysis of genetic sequences, medical diagnosis, and drug discovery. In one striking breakthrough, Alphafold-2, an AI-based approach to predicting 3D protein structure from 1D amino acid

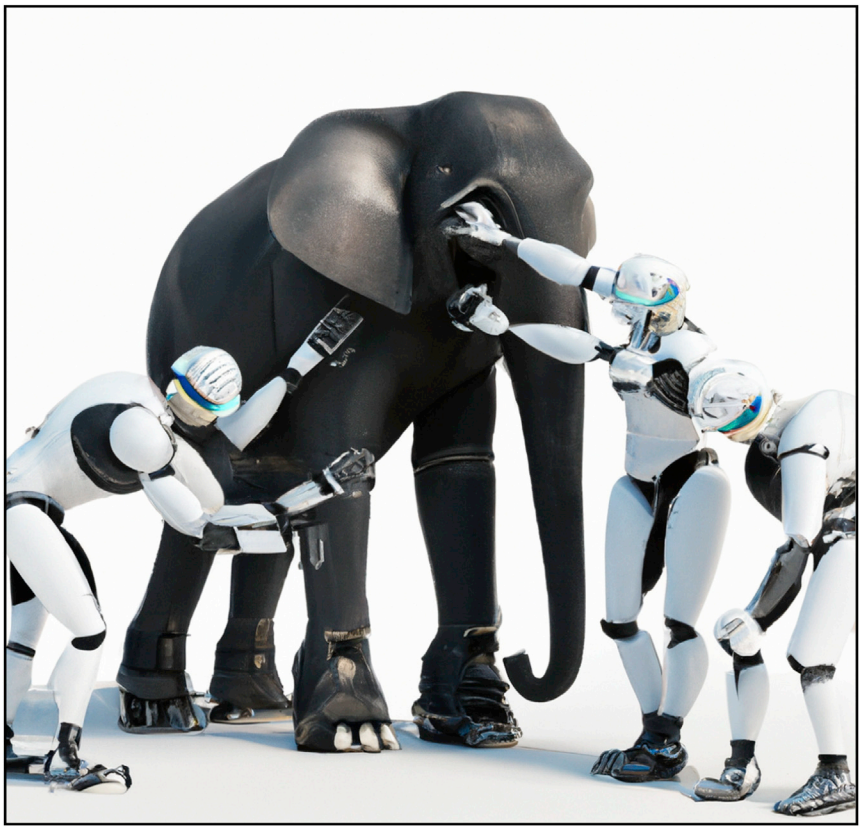

Figure 1. Example of modern AI success
This illustration was generated automatically from a "text-to-image" AI system, Dalle (Ramesh et al., 2022). The input to the system was the textual prompt "Three blind men touching an elephant trying to figure out what it is, except the blind men are humanoid robots."

sequence, recently leap-frogged all previous algorithms in the most recent competition (Jumper et al., 2021). We can expect more such breakthroughs in the coming years, in a wide range of domains.

AI tools for modeling the brain

A second important application of AI involves using ANNs as a model of neural computation. ANNs were originally developed as models of the brain. The pioneering researchers, such as John von Neumann (who invented the modern "von Neumann" computer architecture) and Frank Rosenblatt (who invented the "perceptron," the first neural network system to learn from examples), had as their goal not only to build machines that could mimic human thought and reasoning, but also to understand how the brain computes (Lindsay 2021). This use of ANNs as a model of real neural computation was further pursued in the 1980s by cognitive scientists whose explicit goal was to understand and model human cognition. Similarly, the research in the 1980s that led to the development of reinforcement learning was also primarily concerned with modeling how animals learned from "trial-and-error" (Sutton and Barto 2018). Indeed, at that time, a key driving premise of cognitive science was that we could study intelligence as a general phenomenon, using AI models to understand the mind and, in turn, using our understanding of the mind to build better AI systems. Researchers from various domains, including both those who used ANN or reinforcement learning models and those who used more traditional logic-based models (sometimes referred to as "Good Old-Fashioned AI," or GOFAI), agreed that neuroscience/psychology and AI were concerned with many of the same problems and would benefit from interdisciplinary interactions.

Although the application of AI models to cognitive science and neuroscience fell out of favor in the 1990s and early 2000s, the recent success of "deep learning" has rekindled interest in these approaches over the past decade. Thanks to a combination of much larger computational power, much larger datasets, and some new tweaks to the models, AI researchers were able to engineer ANNs that could finally fulfill their potential. The last decade has seen stunning advances in the ability of AI to solve difficult problems once thought intractable for artificial systems. Throughout this time neuroscientists also began using ANNs in their originally intended purpose, as models of real neural computation (Richards et al., 2019). In part, these recent successes were facilitated by newly developed experimental techniques for monitoring the activity of large populations of neurons. This allowed researchers to compare ANNs and real brains more directly, leading to the discovery that the representations that emerge in ANNs trained on relevant tasks can bear striking resemblance to those seen in the real brain (Richards et al., 2019). This correspondence has emerged in both feedforward and recurrent neural networks across many brain areas, including low- and high-level visual areas, language areas, motor areas, and prefrontal areas (Figure 2A; Yamins et al., 2014). For example, it was found that inferotemporal cortex, a region critical for representing object identity in primates, is spatially organized into a map of object space whose two axes are the same as those in a late layer of the deep network trained on object classification (Bao et al., 2020). At the same time, computational neuroscientists interested in understanding learning and plasticity in the brain began looking at the techniques used to train ANNs and found that some of the same principles could, in theory, be in operation in the brain (Figure 2B; Richards et al., 2019). The result has been a huge increase in research using AI systems to model many aspects of animal behavior and cognition.

There is also a renewed hope that neuroscience will be able to provide additional insights for the development of new ANN approaches that can further advance the state of AI moving forward. There are still many areas where brains clearly excel over ANNs, and these

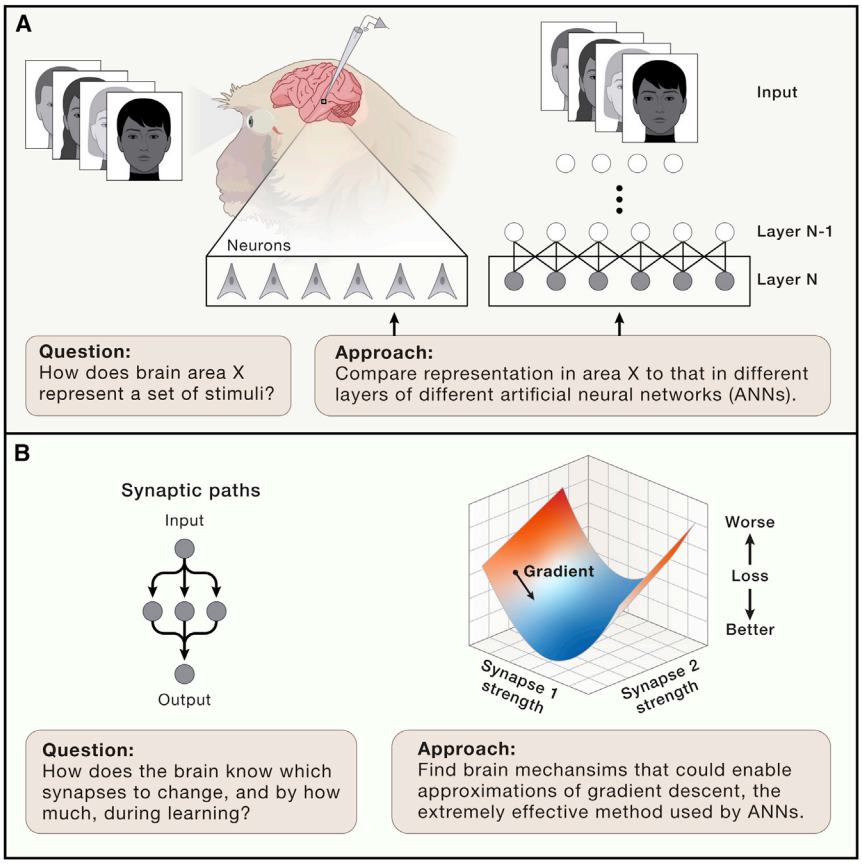

Figure 2. How ANNs are helping us to understand brain function and plasticity

(A) ANNs are providing neuroscientists with a rich set of models to explain neural activity in different brain areas, an important step toward understanding how complex functions such as face recognition are accomplished by the brain. Researchers can show a set of stimuli to an animal, record responses of neurons in that area, and then ask how well different ANNs capture the responses. One popular strategy is to model a neuron as a linear combination of ANN units and then ask how much variance in the neural response is explained by the linear model.

(B) In both brains and ANNs, there are many possible synaptic paths leading from input to output. For example, here there are three possible paths and six possible synapses. This makes it challenging to know which synapses should change during learning, an issue known as the "credit assignment problem." ANNs solve this problem by using the gradient of a loss function to determine how to change synapses. Recent work in computational neuroscience shows the brain has mechanisms that could enable approximations to gradient descent.

provide rich grounds for new, neuro-inspired ANN models. One example is the machine learning problem of catastrophic forgetting, whereby learning new examples causes forgetting of old samples. A recent DARPA grand challenge to solve this problem identified "replay" as the most effective solution (Kudithipudi et al., 2022). This is a biological mechanism in which the hippocampus re-activates already learned memories. Another line of inquiry arises from the recognition that a great deal of animal (including human) behavior is innate and thus somehow embedded in the genome. This has inspired attempts to create ANNs whose structure, like the structure of biological networks, must pass through a "genomic bottleneck" (Koulakov et al., 2021). There is growing recognition that neuroscience has a role to play in guiding future innovations in AI.

Future perspectives

AI is influencing all scientific disciplines by providing new tools to analyze high-dimensional data. In particular, in neuroscience AI is providing powerful new models to explain how brains compute. The use of AI as a model of neural computation can be considered a fulfillment of AI's original birth purpose. Yet AI is also showing itself to be a truly disruptive, paradigm-shifting child, with recent advances compelling neuroscientists to rethink the entire epistemological basis for their enterprise. What does it mean to understand the brain? For much of the history of neuroscience, the answer has been that understanding entails "being able to explain as much neural activity as possible by a simple model." For example, Hubel and Wiesel's model of neurons in primary visual cortex as edge detectors was considered successful because it could explain, with a single parameter model ("edge orientation"), stimulus-evoked responses of neurons to a wide range of different stimuli. However, with the advent of models containing millions of parameters, the notion of "simple explanation" is now becoming fuzzy. What is the value of explaining a brain area by a neural network containing countless units that itself may not be well understood? Many answers are possible, e.g., to enable new predictions (even if a weather model contains millions of parameters, it is still useful if it can accurately predict the weather), to identify the optimal architecture and learning rule used by the brain (these system descriptions are much lower dimensional and hence understandable compared to patterns of weights and resulting stimulus selectivities of units), or to enable important practical applications (e.g., brain machine interfaces and sensory prostheses).

The great physicist Richard Feynman famously said, "I do not understand what I cannot build," but recent successes of AI suggest that "I cannot understand even that which I can build." Even within the field of deep learning there is a strong sense that the models are shockingly, unreasonably effective. The field is currently progressing in an evolutionary mode, in which the most successful models rise to the top not so much due to engineering based on foundational principles but due to survival of the fittest. As a consequence, intense interest is now growing in theoretical machine learning, in hopes of gaining deeper understanding of effective models (for example, one insight that has emerged is that deep networks do not get stuck in local minima because in high dimensions, true local minima are virtually non-existent). For a neuroscientist, this raises the question: what is the

value of overcoming incredible technical challenges to record activity from tiny neurons in a fragile sheet of brain tissue encased in a hard and opaque skull, when we do not even understand the principles of completely transparent deep networks in which the activity of every unit in response to any stimulus/perturbation can be instantly measured? A determined graduate student might answer, "The machines aren't conscious, and I want to understand an intelligence capable of begetting consciousness." But what if one day the machines do aver that they are fully conscious—perhaps even exclaiming to us how intensely they dream about understanding consciousness? Will neuroscientists—at least those motivated by the desire to understand the brain—then put away their microscopes and toss their electrodes?

ACKNOWLEDGMENTS

We would like to acknowledge funding from Schmidt Futures (A.Z.), the Mathers Foundation (A.Z.), the Schwartz Foundation (A.Z.), CIFAR (B.R.), CFREF-HBHL (B.R.), NSERC (B.R.), HHMI (D.T.), NIH RO1EY030650-05 (D.T.), ONR N00014-20-1-2786 (D.T.).

DECLARATION OF INTERESTS

A.Z. consults for and is a founder of Cajal Neuroscience and consults for DVL. B.R. consults for DeepMind Inc.

REFERENCES

Bao, P., She, L., McGill, M., and Tsao, D.Y. (2020). A Map of Object Space in Primate Inferotemporal Cortex. Nature *583*, 103–108. https://doi.org/10.1038/s41586-020-2350-5.

Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. Neuron *95*, 245–258. https://doi.org/10.1016/j.neuron.2017.06.011.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Koulakov, A., Shuvaev, S., Lachi, D., and Zador, A. (2021). Encoding Innate Ability through a Genomic Bottleneck. https://doi.org/10.1101/2021.03.16.435261.

Kriegeskorte, N., and Douglas, P.K. (2018). Cognitive Computational Neuroscience. Nat. Neurosci. *21*, 1148–1160. https://doi.org/10.1038/s41593-018-0210-5.

Kudithipudi, D., Aguilar-Simon, M., Babb, J., Bazhenov, M., Blackiston, D., Bongard, J., Brna, A.P., Raja, S.C., Cheney, N., Clune, J., et al. (2022). Biological Underpinnings for Lifelong Learning Machines. Nature Machine Intelligence *4*, 196–210. https://doi.org/10.1038/s42256-022-00452-0.

Lindsay, G. (2021). Models of the Mind: How Physics, Engineering and Mathematics Have Shaped Our Understanding of the Brain (Bloomsbury Publishing).

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning. Nat. Neurosci. *21*, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y.

MICrONS Consortium, Alexander Bae, J., Baptiste, M., Bodor, A.L., Brittain, D., Buchanan, J., Castro, M.A., Bumbarger, D.J., Celii, B., Cobos, E., Collman, F., et al. (2021). Functional Connectomics Spanning Multiple Areas of Mouse Visual Cortex. Preprint at bioRxiv. https://doi.org/10.1101/2021.07.28.454025.

Rajpurkar, P., Chen, E., Banerjee, O., and Topol, Eric J. (2022). AI in Health and Medicine. Nat. Med. *28*, 31–38. https://doi.org/10.1038/s41591-021-01614-0.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. Preprint at arXiv. https://doi.org/10.48550/arXiv.2204.06125.

Richards, B.A., Lillicrap, T.P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R.P., de Berker, A., Ganguli, S., et al. (2019). A Deep Learning Framework for Neuroscience. Nat. Neurosci. *22*, 1761–1770. https://doi.org/10.1038/s41593-019-0520-2.

Sapoval, N., Aghazadeh, A., Nute, M.G., Antunes, D.A., Balaji, A., Baraniuk, R., Barberan, C.J., Dannenfelser, R., Dun, C., Edrisi, M., et al. (2022). Current Progress and Open Challenges for Applying Deep Learning across the Biosciences. Nat. Commun. *13*, 1728. https://doi.org/10.1038/s41467-022-29268-7.

Sutton, R.S., and Barto, A.G. (2018). Reinforcement Learning: An Introduction. Second edition. Adaptive Computation and Machine Learning Series (Cambridge, Massachusetts: The MIT Press).

Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-Optimized Hierarchical Models Predict Neural Responses in Higher Visual Cortex. Proc. Natl. Acad. Sci. USA *111*, 8619–8624. https://doi.org/10.1073/pnas.1403112111.