Artificial Intelligence: A Brief Introduction for Non-Experts on the Technological Advances That Are Bringing Smart Devices into Our Lives

Friedemann Zenke[1,2] and Koshika Yadava

[1] Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland.
[2] Faculty of Science, University of Basel, Basel, Switzerland.

Published as: Zenke, F., Yadava, K., 2024. A Brief Introduction for Non-Experts on the Technological Advances That Are Bringing Smart Devices into Our Lives, in: Robots and Gadgets: Aging at Home. Les Presses de l'Université Laval. https://doi.org/10.2307/jj.9421088.4

We navigate the world guided by the invisible. As we search the web, Google learns from our behaviours and predicts what we are looking for even before we completely type it. We can even forgo typing and gently whisper our wishes to our smartphones. An amiable virtual assistant accompanies us on our commute and helps us plan our day. At home, smart systems fine-tune lights, music volume, and heating to match our moods across the day and seasons. Artificial intelligence (AI) empowers these smart systems to find patterns within the expanses of our behavioural data to predict and adapt to our ever-changing desires. Even our robot vacuum cleaner constitutes a form of embodied albeit different artificial intelligence, precisely navigating around furniture to grab dirt. At the same time, we converse with algorithms that know how to entertain us. AI is employed even in the bureaucratic humdrum: facial-recognition systems at airports compare our faces to the biometric image stored on our passports and wave us through or redirect us to a human customs officer if further inquiry is needed. Smart technologies rely on our interactions to become "smarter," and our increasing engagement with and dependence on these technologies signal two inevitabilities. On the one hand, the emergence of more mature and competent AI. And on the other hand, an ever-growing presence of these technologies in virtually all aspects of our daily lives. Yet most users of these smart technologies seldom pay much attention to what AI is, how it works, and its similarities to or differences from human intelligence.

Let us first define what intelligence is. Intelligence combines perceiving and retaining information with the ability to apply it in different contexts. For instance, most of us get our food at the local store. Still, when shopping for cheese in a new city, we expect to find it within the dairy section and not in the laundry-detergent section. This expectation arises because we have learned that cheese is milk-based and usually needs to be refrigerated and is thus typically in the corner of the shop with the coolers. Acquiring knowledge through experience allows us to find our bearings in new situations quickly.

This ability is what computer scientists call generalization. Humans and some animal species such as monkeys or crows are very good at it and are remarkably adept at applying prior knowledge to new situations. AI comprises artificial systems capable of solving tasks that previously required the biological intelligence exhibited by humans and animals. Though several different definitions for AI exist, most of them centre on the notions of problem-solving through reasoning and of planning by considering previous knowledge and adapting it from one scenario to a new, unfamiliar setting (Russell and Norvig 2020). In other words, one can think of intelligence as a meta-skill that allows one to solve novel problems, acquire new skills, and learn.

Researchers further distinguish between narrow and general AI. Most current AI systems fall into the former category. For instance, a self-driving car can keep within its lane while on a highway, and a robot vacuum cleaner will move around objects in your home while cleaning. Yet neither would apply their driving skill to the other domain nor learn to play chess. Humans, on the other hand, are capable of mastering diverse tasks. The machine equivalent of such intelligence would be what we call artificial general intelligence (AGI). Although AGI is the object of intense research efforts backed by major industries, we are presumably still far from developing true AGI. In the following, we focus on narrow-AI systems whose possibilities are limited to one or several specific tasks, such as speech recognition and computer vision. Their capabilities are still impressive, and researchers have made great strides in developing them further.

Recent advances and current possibilities of AI

The first Grand Challenge for autonomous vehicles, a 142-mile course through the Californian desert, was organized by the Defence Advanced Research Projects Agency (DARPA) in 2004. No vehicle completed the course (DARPA, 2004). However, only a few years later, in 2009, a self-driving vehicle from Google had completed over ten drives that were longer than one hundred miles each. Self-driving cars use advanced vision sensors, radar, and lidar to constantly map their surroundings while in motion. Machine-learning algorithms process these data streams and infer the position of the vehicle with respect to the lane on the road, other vehicles, and any unforeseen obstacles. Based on this information, they control actuators to keep the car in lane and to avoid collisions. When the algorithms encounter a situation they cannot handle, they usually pass control back to the human driver or initiate an emergency stop of the vehicle.

Driving and many tasks we perform as humans rely on vision and, consequently, improving computer vision algorithms is an intense focus area in AI research. The annual ImageNet Large Scale Visual Recognition Challenges (Russakovsky et al. 2015) have showcased just how far this field has progressed. At the competition, computer vision algorithms have to categorize an object or animal depicted in a photograph. While the categories are known in advance, the competing algorithms have never before seen the images that they have to classify. For instance, the algorithms would have to “know” whether a picture they have never seen before shows a chair, a container ship, or a cow. In 2012, AlexNet, a convolutional neural network (Krizhevsky et al. 2012), with an architecture inspired by the primate visual system, won the ImageNet competition by a large margin. AlexNet’s success largely relied on a novel, “deep” neural-network architecture, and their win heralded the age of deep neural networks, which have since transformed most modern AI approaches. We will learn more about what deep neural networks are in the next section but, for now, let us focus on what they can do.

DeepMind, a deep learning start-up that has since become a subsidiary of Google, developed an AI software called AlphaGo that plays the game of Go. In 2016, AlphaGo, primarily built on progress in deep-neural-network technology, won in the game of Go against a human grandmaster (Silver et al. 2016). It was an extraordinary feat. In chess, a computer had beaten the champion, Garry Kasparov, twenty years earlier, in 1997, by simply using brute force to simulate a myriad of possible games to assess its next move. This strategy is not viable in Go because there are too many possible games, and using it would lead to a combinatorial explosion. There are too many games for even the most powerful supercomputers to simulate. Modern AI systems have parted with purely brute-force strategies and instead take inspiration from the brain, which uses neural networks trained through experience to detect and evaluate patterns. This knowledge enables neural networks to, for instance, judge one chess position resulting from a move as “looking” more advantageous than another, without necessarily having to play out all possible future games that could ensue. And, presumably, this strategy is more like the intuitions that grandmasters like Kasparov, through extensive training, have learned to use to excel at their trade.

The breakthroughs with deep neural networks keep coming. In 2020, OpenAI released GPT-3, an AI system with the ability to write and talk almost like a human about virtually any topic you could think of. But AI has not been limited to self-driving cars, playing games, and robotic banter. In 2020 as well, DeepMind released AlphaFold-2, an AI system (Jumper et al. 2021) that constitutes a chemistry Nobel Prize–worthy advance in protein-structure prediction and which has the potential to fundamentally change pharmaceutical research. Simultaneously, there has been tremendous progress for AI in health-care and nursing applications. For instance, AI is helping to identify, classify, and quantify pathologies in medical images (Shen et al. 2017). Similarly, in elderly care, AI is heralding a new age by providing smart systems that monitor human movements through unobtrusive computer vision systems and alert caregivers in the event of falls or when they detect behaviour that is out of the ordinary (Corbyn, 2021). AI is also advancing the capabilities of neuroprosthetics, thereby allowing paralyzed persons with anarthria to speak (Moses et al. 2021). To that end, an implanted electrode array picks up the person's brain signals from the speech sensorimotor cortex. These signals are subsequently processed by a neural processing system that infers the probabilities of the person thinking of specific words that they want to communicate. Finally, the predicted word probabilities are fed into a language model similar to GPT-3, and the outputs are decoded as sentences displayed on a screen or enunciated through an artificial speech-generation system. So how do modern AI systems achieve this level of accomplishment?

The technical and conceptual progress that has enabled modern AI technologies

When George Boole 1854) published his *Investigation of the Laws of Thought* some 150 years ago, he believed that one day we would understand the mind and, ultimately, intelligence through logic and mathematical equations that can be written down succinctly. Boole's influential work heralded the information age, laid the theoretical foundations of modern computers, and has defined in large part how people are trying to build AI. It was only around the turn of the new millennium when AI researchers gradually realized that the rules of intelligence seem too complex to be written down by hand.

This realization formed the foundations of machine learning, which is concerned with "learning" the rules and identifying the relevant patterns from the data. Deep artificial neural networks (ANNs), which constitute one specific branch of machine learning, proved particularly suitable for this purpose (Krizhevsky et al. 2012; Rumelhart et al. 1986; LeCun et al. 2015; Schmidhuber 2015; Goodfellow et al. 2017). Their design was inspired by that of the brain in which large numbers of identical neurons connect to vast networks whose computational abilities are primarily determined by how they are connected. The central technological advance that made deep learning possible was that researchers have worked out effective algorithms that can learn from large amounts of training data and automatically adjust the myriad of synaptic connections, thereby creating ANNs that can complete a particular task such as recognizing what is in an image (Figure 1.1).

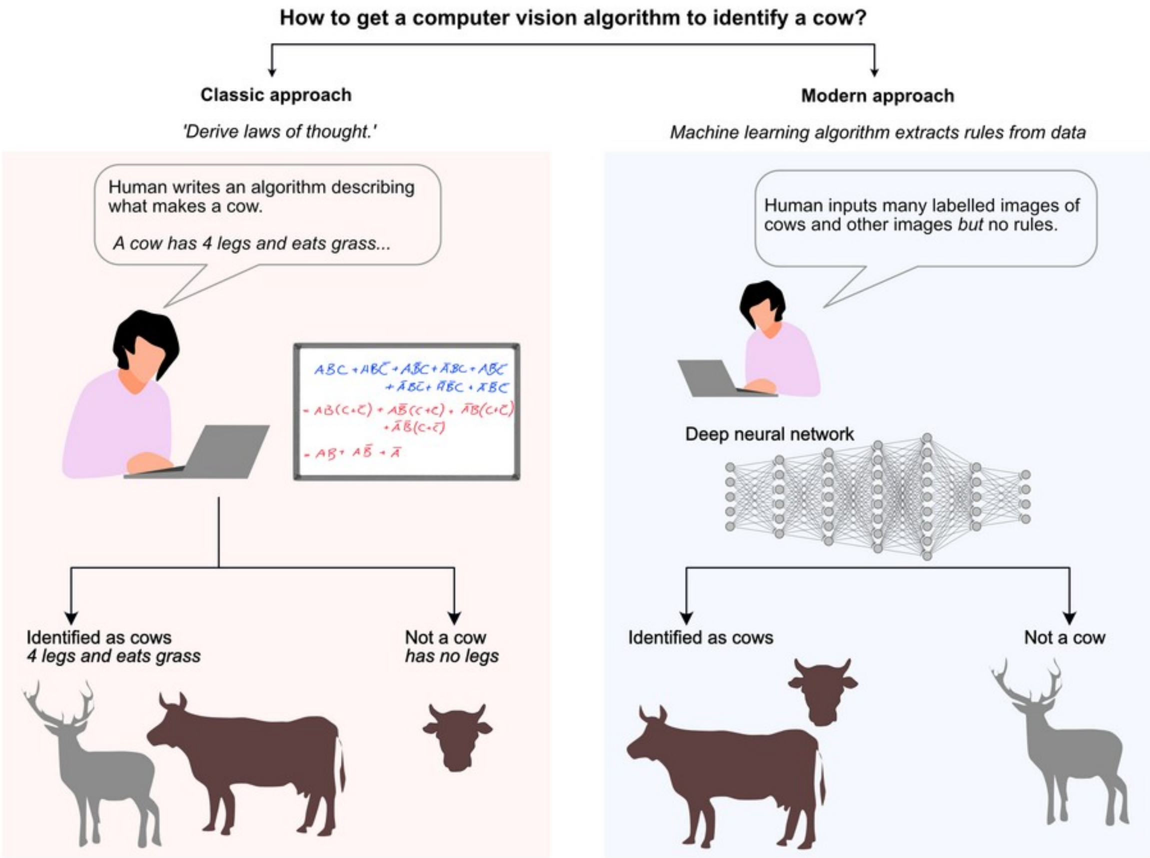

Figure 1.1: Classic and Current Approaches to Identifying an Image

Architecturally, ANNs need to be deep, which means that information travels through many layers before an output is generated. Moreover, depending on the task at hand, neurons need to be pre-wired in specific ways. For instance, in convolutional neural networks, which are essential for computer vision, each neuron in the input layers receives input from only a small part of the image. Then, neurons in subsequent layers receive input from neighbouring neurons in the previous layer, thereby gradually pooling over local information in the input image as the data flow through the ever-deeper layers. To train these deep networks, computer scientists had to advance their theoretical understanding of network dynamics to allow for effective weight-initialization strategies and develop new training algorithms to update the many connection weights more efficiently. Another advance was the emergence of powerful computational capabilities, which arrived in the form of graphic processing units (GPUs), initially developed for the gaming industry. Unlike the highly specialized central processing units (CPUs) at the heart of every computer, GPUs perform simpler operations but allow for massive parallelism. This parallelism allows the application of the same algorithmic functions to different data in parallel, which proved ideal for speeding the vast neural-network simulations used in AI, which require that many identical neurons receive distinct inputs. Thus, GPUs have increased sheer computing power. There now exist diverse specialized hardware tailored for deeplearning applications. Finally, and perhaps most importantly, training deep neural networks requires large amounts of data from which synaptic connections are “learned.” The underlying optimization algorithms are said to operate end to end in that they directly distill knowledge from vast amounts of data into the connections between the deep network layers. That is ultimately the reason why we speak of “deep” learning.

While its reliance on learning from data is deep learning's strength, it is also one of its potential weaknesses because nothing prevents implicit or explicit biases present in the data from being directly absorbed by the AI that relies on it, thus, AI systems can develop racial, gender, age, and more complex biases (Buolamwini and Gebru 2018; Bender et al. 2021). While these biases may be hard to detect, they really impact people's lives. Today, AI systems are involved in hiring decisions, facial recognition at airports, and detection of credit-card fraud, to name only a few examples. As AI's influence on our everyday lives grows, so does the adverse effect of any biases in the datasets used for training them. Therefore, avoiding unwanted biases is essential for any current or future AI system.

Major hurdles on the way to ubiquitous AI

As AIs mature and become more intelligent, the potential applications of smart devices in all walks of life seem to be growing exponentially. At the same time, the extensive computational resources required by modern AI systems are increasing (Lasse et al. 2020), thereby creating a significant impediment to widespread deployment of intelligent devices in, for instance, mobile and smart-home applications. One major issue is the energy cost of running continuous inference on frame-based video or audio streams. Consequently, when you verbally ask your digital personal assistant or you speak instead of typing into your smartphone, the speech audio is not processed on your phone but is instead sent to a cloud server for automatic speech recognition. This transfer reduces the computational burden on your phone, thereby allowing it to operate with fewer hardware requirements and extending its battery life. However, this reliance on a cloud server for processing creates several issues. First, it requires a permanent internet connection. Thus, when driving through an area with a poor mobile connection, the service might not be available. Second, it means that your data are sent to the cloud, with the obvious privacy implications. Thirdly, the transfer introduces a communication delay. This delay may not be an issue when you tell your smart-home speaker to dim the lights. Still, when considering time-critical applications such as driving decisions in an autonomous vehicle, you would want to avoid any delay.

Curiously, a standard GPU used for deep learning applications consumes several hundreds of watts while the human brain, which avoids all the above issues, runs flawlessly on a comparatively limited power budget of approximately twenty-five watts. This discrepancy has inspired decades of research in neuromorphic engineering (Mead 1990; Mead and Ismail 2012; Indiveri et al. 2011; Schuman et al. 2017; Boahen 2017). Schuman et al. (2017) tried to achieve similar efficiency, taking inspiration from the brain by copying its essential architectural features, some of which differ fundamentally from the neural network architectures used in modern AI systems.

Brain-inspired neuromorphic technologies for pervasive AI

While deep learning leverages ANNs, which are intrinsically biological in inspiration, such artificial networks are simulated on processors and GPUs that serve as the computational substrate and have a circuit architecture that is fundamentally different from that of the brain. In broad strokes, neuromorphic engineering has taken on the challenge of building computational substrates in electronics that emulate neural network computation instead of simulating it. Intense research efforts in this direction have resulted in an increasing number of neuromorphic substrates being available today (Indiveri et al. 2011; Merolla et al. 2014; Davies et al. 2018; Grübl et al. 2020) that allow for an efficient emulation of brain-inspired neural networks. There are two major design characteristics that many neuromorphic substrates try to copy from neurobiology.

The first is that the brain uses in-memory computation, whereby synapses are the physical location of long-term information storage and also of the elements that carry out the computation during inference, that is, where the network processes information. This co-location of memory and computation is fundamentally different from conventional computers, which generally adhere to von Neumann architecture, in which memory and CPU are separated. In practice, this separation creates a computational bottleneck for neural networks since it requires a constant back-and-forth of synaptic-weight values between CPU and memory, thereby resulting in excessive power consumption. Significant research efforts, therefore, are focusing on building in-memory, compute-enabled substrates based on novel device technologies such as memristors (Chua 1971; Strukov et al. 2008; Marković et al. 2020; Joshi et al. 2020; Yao et al. 2020), resistors that change their resistance in an activity-dependent manner.

The second characteristic of many neuromorphic substrates is spiking neurons that closely mimic the signalling properties of biological neurons in the brain. Conventional ANNs used in deep learning rely on analogue neuronal activation functions, which allows neuronal output to take on a real-valued quantity. Biological neurons, however, communicate through action potentials or spikes, which are all-or-nothing events localized in time. In other words, when a spiking neuron does not “fire,” no information is communicated. This contrasts with conventional ANN implementations in which even an output value of zero is communicated. From a power-efficiency perspective, the advantage of spiking seems obvious: neurons only communicate with other neurons when they have something important to say. Neurobiology appears to make extensive use of this power-saving mechanism in that spiking activity in many brain areas is exceptionally sparse, a property commonly ascribed to the superior power efficiency of the brain (Sterling and Laughlin 2015; Davidson and Furber 2021). Based on this idea, researchers have developed neuromorphic digital vision sensors, the brain-inspired equivalent of pixel-based cameras, that only generate spikes when changes in a visual scene are detected (Lichtsteiner et al. 2008). For instance, a pixel-based surveillance camera generates a constant data rate of twenty-five frames per second, requiring downstream processing that consumes energy. In contrast, a neuromorphic vision sensor only transmits spikes when the scene changes (for instance, when somebody moves in the sensor’s field of view), thereby curbing the data rate, which provides tremendous potential for power savings, provided that a suitable event-based processing system is used.

While power-efficient neuromorphic substrates could provide such processing, given that they had been available for some time, one major challenge remained. Like an ANN, the spiking neural network (SNN) that operates a spiking neuromorphic substrate requires training. However, practical training algorithms that can simultaneously deal with spiking and run on neuromorphic hardware were lacking until recently, thereby creating a performance gap between deep neural networks and SNNs. Luckily, recent advances in training algorithms for SNNs seem capable of closing this gap, thereby paving the way for training power-efficient deep spiking-network models on neuromorphic hardware (Pfeiffer & Pfeil 2018; Shrestha and Orchard 2018; Wu et al. 2018; Neftci et al. 2019; Bellec et al. 2020; Kaiser et al. 2020; Bohnstingl et al. 2020; Büchel et al. 2021).

These advances open the door for a diversity of exciting applications. First, the potential of emulating SNNs on ultra-low-power neuromorphic hardware (Blouw et al. 2019; Bojian et al. 2021; Cramer et al. 2022) makes them ideally suited for mobile and always-on applications, for instance, home-surveillance tasks or situations that require a portable device running on a battery. Moreover, their ability to process sparse sensory data with low latency (meaning that the processing does not induce long delays) suggests potential roles for spiking neuromorphic hardware in automotive applications. Further, integrating neuromorphic SNN chips into smart devices will remove the need for communicating with cloud-based servers and any potential associated implications for privacy. Finally, spiking neuromorphic hardware holds the promise of building smart neuroprosthetics that interface directly with biological neural tissue.

What do these advances bring to the future of smart-home technologies? For example, in the context of elderly care, SNN chips will enable always-on speech recognition on battery-powered devices that allow users with reduced vision or mobility to communicate with intelligent devices directly using speech. Simultaneously, the power efficiency of SNN-based solutions holds the promise of vastly expanding battery life, a safety-critical feature, especially for forgetful users.

Similarly, event-based vision sensors deployed around the home will conceivably monitor the inhabitants' every step, detecting a fall or other medical emergency immediately while being able to communicate and call for help if the person is not responsive. Crucially, during normal operation, on-chip processing in the vision sensors themselves, combined with ultra-low-power edge processing, will ensure that no private data are sent to the cloud. All data remain on site with the user, and external communications are only established when there is a genuine problem requiring intervention.

Conclusion

In summary, conceptual advances in deep learning combined with unprecedented amounts of data and computational power have propelled AI to an unparalleled level of ability, thereby opening the door to smart devices. At the same time, novel training techniques for SNNs are allowing the unlocking of the unprecedented energy efficiency of neuromorphic hardware technologies. Together, these technical advances will make it possible for AI to accompany us in our everyday lives, which holds incredible opportunities. Nonetheless, important concerns regarding privacy and the ethics of their use remain.

References

Bellec, G. et al. 2020. "A Solution to the Learning Dilemma for Recurrent Networks of Spiking Neurons." *Nature Communications* 11: 3625. https://doi.org/10.1038/s41467-020-17236-y.

Bender, E. M. et al. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *FAccT '21*: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. New York: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Boahen, K. (2017). "A Neuromorph's Prospectus." *Computing in Science Engineering* 19 (2): 14–28. https://doi.org/10.1109/MCSE.2017.33

Boole, G. (1854). *An Investigation of the Laws of Thought*. London: Walton and Maberly.

Bohnstingl, T. et al. (2020). "Online Spatio-Temporal Learning in Deep Neural Networks. IEEE Transactions on Neural Networks and Learning Systems 1–15. https://doi.org/10.1109/TNNLS.2022.3153985.

Bojian, Y. et al. (2021). "Accurate and Efficient Time-Domain Classification with Adaptive Spiking Recurrent Neural Networks." *Nature Machine Intelligence*, 3 (10): 905–13. https://doi.org/10.1038/s42256-021-00397-w.

Blouw, P. et al. (2019). “Benchmarking Keyword Spotting Efficiency on Neuromorphic Hardware.” In *Proceedings of the 7th Annual Neuro-Inspired Computational Elements Workshop*, 1. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3320288.3320304

Büchel, J. et al. (2021). “Supervised training of spiking neural networks for robust deployment on mixed-signal neuromorphic processors”. Scientific Reports 11, 23376. https://doi.org/10.1038/s41598-021-02779-x

Buolamwini, J., and T. Gebru. (2018). “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, edited by S. A. Friedler and C. Wilson, 77–91. Proceedings of Machine Learning Research.

Chua, L. (1971). “Memristor – The Missing Circuit Element.” *IEEE Transactions on Circuit Theory*, 18 (5): 507–19. https://doi.org/10.1109/TCT.1971.1083337.

Corbyn, Z. (2021). “The Future of Elder Care Is Here – And It’s Artificial Intelligence.” *The Guardian,* 3 Jun, 2021.

Cramer, B. et al. (2022). “Surrogate Gradients for Analog Neuromorphic Computing.” *Proceedings of the National Academy of Sciences* 119 (4): e2109194119. https://doi.org/10.1073/pnas.2109194119.

DARPA (Defence Advanced Research Projects Agency). 2004.. *Grand Challenge 2004*: *Final Report*: *Technical Report*. Arlington, VA: DARPA. https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/DARPA/15-F-0059_GC_2004_FINAL_RPT_7-30-2004.pdf.

Davidson, S., and S. B. Furber. 2021. “Comparison of Artificial and Spiking Neural Networks on Digital Hardware.” *Frontiers in Neuroscience*, 15. https://doi.org/10.3389/fnins.2021.651141.

Davies, M. et al. 2018. “Loihi: A Neuromorphic Manycore Processor with On-Chip Learning.” *IEEE Micro* IEEE Micro 38, 82–99. https://doi.org/10.1109/MM.2018.112130359.

Goodfellow, I. et al. 2017. *Deep Learning*. Cambridge, MA: MIT Press.

Grübl, A. et al. 2020. “Verification and Design Methods for the BrainScaleS Neuromorphic Hardware System.” *Journal of Signal Processing Systems*, 92 (11): 1277–92. https://doi.org/10.1007/s11265-020-01558-7.

Indiveri, G. et al. 2011. “Neuromorphic Silicon Neuron Circuits.” *Frontiers in Neuroscience* 5: 73. https://doi.org/10.3389/fnins.2011.00073.

Joshi, V. et al. 2020. “Accurate Deep Neural Network Inference Using Computational Phase-Change Memory.” *Nature Communications* 11 (1): 2473. https://doi.org/10.1038/s41467-020-16108-9.

Jumper, J. et al. 2021. “Highly Accurate Protein Structure Prediction with AlphaFold.” *Nature*, 596 (7873): 583–9. https://doi.org/10.1038/s41586-021-03819-2.

Kaiser, J. et al. 2020. “Synaptic Plasticity Dynamics for Deep Continuous Local Learning (DECOLLE).” *Frontiers in Neuroscience*, 14. https://doi.org/10.3389/fnins.2020.00424.

Krizhevsky, A. et al. 2012. “ImageNet Classification with Deep Convolutional Neural Networks.” In *Advances in Neural Information Processing Systems* Vol. 25, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 1097–1105. Red Hook, NY: Curran Associates.

Lasse, F. et al. 2020. “Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models.” *arXiv:*2007.03051.

LeCun, Y. et al. 2015. “Deep Learning.” *Nature*, 521 (7553): 436–44. https://doi.org/10.1038/nature14539.

Lichtsteiner, P. et al. 2008. “A 128x128 120 dB 15us Latency Asynchronous Temporal Contrast Vision Sensor.” *IEEE Journal of Solid-State Circuits*, 43 (2): 566–76. https://doi.org/10.1109/JSSC.2007.914337.

Marković, D. et al. 2020. “Physics for Neuromorphic Computing.” *Nature Reviews Physics* 2 (9), 499–510. https://doi.org/10.1038/s42254-020-0208-2.

Mead, C. 1990. “Neuromorphic Electronic Systems.” In *Proceedings of the IEEE*, 78 (10): 1629–36. https://doi.org/10.1109/5.58356.

Mead, C., and M. Ismail., eds. 2012. *Analog VLSI Implementation of Neural Systems*. Norwell, MA:: : Kluwer Academic Publishers.

Merolla, P. A. et al. 2014. “A Million Spiking-Neuron Integrated Circuit with a Scalable Communication Network and Interface.” *Science*, 345 (6197): 668–73. https://doi.org/10.1126/science.1254642

Moses, D. A. et al. 2021. “Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria.” *New England Journal of Medicine* 385 (3): 217–27. https://doi.org/10.1056/NEJMoa2027540.

Neftci, E. O. et al. 2019. “Surrogate Gradient Learning in Spiking Neural Networks: Bringing the Power of Gradient-Based Optimization to Spiking Neural Networks.” *IEEE Signal Processing Magazine*, 36 (6): 51–63. https://doi.org/10.1109/MSP.2019.2931595.

Pfeiffer, M., and T. Pfeil. 2018. “Deep Learning with Spiking Neurons: Opportunities and Challenges.” *Frontiers in Neuroscience*, 12. https://doi.org/10.3389/fnins.2018.00774.

Rumelhart, D. E. et al. 1986. “Learning Representations by Back-Propagating Errors.” *Nature*, 323 (6088): 533–6. https://doi.org/10.1038/323533a0.

Russakovsky, O. et al. 2015. “ImageNet Large Scale Visual Recognition Challenge.” *International Journal of Computer Vision* 115 (3): 211–52. https://doi.org/10.1007/s11263-015-0816-y.

Russell, S., and P. Norvig. 2020. *Artificial Intelligence: A Modern Approach*. 4th ed. Hoboken: Pearson.

Schmidhuber, J. 2015. “Deep Learning in Neural Networks: An Overview.” *Neural Networks* 61: 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Schuman, C. D. et al. 2017. “A Survey of Neuromorphic Computing and Neural Networks. in Hardware.” *arXiv* 1705.06963.

Shen, D et al. 2017. “Deep Learning in Medical Image Analysis.” In *Annual Review of Biomedical Engineering*, 19 (1): 221–48. https://doi.org/10.1146/annurev-bioeng-071516-044442.

Shrestha, S. B., and G. Orchard. 2018. “SLAYER: Spike Layer Error Reassignment in Time,” in *Advances in Neural Information Processing Systems*, edited by S. Bengio et al., 1419–28. Red Hook, NY: Curran Associates.

Silver, D. et al. 2016. “Mastering the Game of Go with Deep Neural Networks and Tree Search.” *Nature*, 529 (7587): 484–9. https://doi.org/10.1038/nature16961.

Sterling, P., and S. Laughlin. 2015. *Principles of Neural Design*. Cambridge, MA: MIT Press.

Strukov, D. B. et al. 2008. “The Missing Memristor Found.” *Nature*, 453 (7191): 80–83. https://doi.org/10.1038/nature06932.

Yao, P. et al. 2020. “Fully Hardware-Implemented Memristor Convolutional Neural Network.” *Nature*, 577 (7792): 641–46. https://doi.org/10.1038/s41586-020-1942-4.

Wu, Y. et al. (2018). “Spatio-Temporal Backpropagation for Training High-Performance Spiking Neural Networks.” *Frontiers in Neuroscience*, 12: 331. https://doi.org/10.3389/fnins.2018.00331

Glossary of Technical Terms

Deep learning, a subdiscipline of machine learning that uses deep artificial-neural-network models with many layers, essentially underlies the success of artificial intelligence. The term “learning” implies that the connections in deep neural networks are fine-tuned algorithmically using large datasets of labelled examples, for instance, images with known content. One speaks of “training” a deep neural network.

Convolutional neural networks are a specific type of deep neural network with a local connectivity structure that is commonly used in computer vision to identify objects. Convolutional neural networks are inspired by the visual system in humans and other primates, whereby neurons in lower layers only see a small part of an image. Subsequent layers pool information from increasing fractions of the picture. This biologically inspired connectivity structure is vital for the ability of such networks to generalize to unseen images.

Von Neumann architecture: This classic computer architecture posits a physical separation between a central processing unit and memory. It underlies virtually all modern computers but is notably different from biological neural networks in the brain, where computing elements (neurons) and memory (synapses) are co-located in the brain tissue.

Neuromorphic computing substrate: Neuromorphic substrates are computing architectures inspired by biological neural networks. Neuromorphic substrates typically aim to co-locate computational elements and memory to avoid bottlenecks resulting from their separation, as with the von Neumann architecture. The primary motivation for this co-location is to produce more power-efficient and fault-tolerant computing systems that are more like biological brains.

Edge computing: Edge computing is a distributed-computing paradigm that brings computation closer to the sensors that produce the data. For instance, it can be performed by a small server at home, which contrasts with cloud computing, in which computation occurs on servers in the cloud that may be physically far away. Edge computing decreases response times, saves communication bandwidth, and improves privacy.

Action potential or spikes: Biological neurons communicate with short electrical pulses called action potentials or spikes. This procedure contrasts with the neuronal-activation function used by conventional deep neural networks, which transmit analogue-valued outputs. Spiking communication presumably plays a significant role in the unparalleled power efficiency of biological brains as compared to the deep neural networks used in machine learning.