frontiers | Frontiers for Young Minds

ENGINEERING AND TECHNOLOGY
Published: 07 May 2024
doi: 10.3389/frym.2024.1164958

WILL LEARNING MACHINES TAKE OVER THE WORLD?

Yann LeCun [1,2]*

[1] *Courant Institute, New York University (NYU), New York, NY, United States*

[2] *Meta, AI-FAIR, New York, NY, United States*

YOUNG REVIEWERS:

OISIN
AGE: 12

ZI-AN
AGE: 8

Learning is an integral part of our lives, and the lives of all animals, but do you realize how wondrous our learning ability is? When we try to build machines that can learn, we are faced with deep questions about the nature and functioning of intelligence. In this article, I will tell you about special artificial networks, called neural networks, that mimic the brain to produce intelligent behavior. Neural networks are an integral part of artificial intelligence, widely used in many daily applications, from face recognition to autonomous driving. I am certain that neural networks will play a key role in our future lives, making them more comfortable and improving our understanding of big riddles such as what intelligence is and how our brains work.

Professor Yann LeCun won the Turing Award in 2018, jointly with Prof. Geoffrey Hinton and Prof. Yoshua Bengio, for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.

ARE MACHINES SMARTER THAN HUMANS?

Are you intelligent? I assume that all of you would say "yes". Now, what about your dog or cat? And how about your smartphone? Though intelligence is intuitive for us, it is actually a complex phenomenon [1] and there is no clear definition of intelligence and no absolute measure of how intelligent a system is. What we often do in **artificial intelligence (AI)** is try to reproduce, or surpass, the capabilities of humans and animals, particularly intellectual abilities like understanding language and reasoning.

ARTIFICIAL INTELLIGENCE (AI)

The ability to reproduce, and even surpass, with machines the capabilities of humans and animals.

AI is a moving target. Whenever some problem is solved so that machines outperform humans in a specific task, it is no longer AI. Many things which used to be part of AI are not any more, such as using GPS to navigate the best route or performing complex arithmetical calculations. Through advancing AI we also learned that humans are relatively good, or intelligent, at certain things but pretty bad at other things. For example, until recently, people thought that we were unbeatable at board games like Chess and Go. But today we know that advanced AI systems can beat the best human players—see this documentary to learn about the first AI system (AlphaGo) to beat a human world champion, in 2015. Today, there are other tasks which AI systems can perform better than humans, such as instantaneous translation and image recognition. But even the most intelligent systems are highly specialized: they are good only at a very narrow range of tasks [2].

Truly intelligent machines must be able to do what all animals do—learn [3]. Learning means adapting, or using knowledge gained in experiences to perform well in a task within new scenarios. Humans and animals learn naturally, but how do machines learn? That is the research of **machine learning (ML)**. In ML, we develop systems and **algorithms** that take in data and use it to continuously update their behavior and improve their performance. ML is the basis of many applications, like automatic translation of webpages to different languages, voice control and face recognition in smartphones, driving autonomous cars and analyzing medical data. But the best learning machines still can not learn like humans. For example, it takes humans only about 20–30 h to learn how to drive properly, but currently no machine can drive autonomously as well as a human, even after thousands of hours. A big ML breakthrough was when we figured out how to imitate the way the brain learns.

MACHINE LEARNING (ML)

Teaching machines how to learn, or improve their performance, as result of previous experience, with as little human intervention as possible.

ALGORITHM

A set of instructions that tells machines how to operate.

HOW DO WE LEARN?

Our brains are made of about 100 billion (hundred thousand million) **neurons**, each connected to about 10,000 other neurons (this means we have about 1,000 trillion connections in our brain!). They communicate via electrical signals that they transmit and receive.

NEURONS

Cells in the brain that process information using electrical signals.

The strength of electrical signal each neuron receives from other neurons is determined by the **synapse**—the junction that connects it with other neurons (Figure 1A; to learn more about neurons and synapses, see this paper). When we learn something, the strengths and locations of synapses in our brain change—think of learning as a network of units that receive inputs from other units, perform a computation, and transfer and output via junctions whose strengths of connections change over time. Using these principles, we can create a network of **artificial neurons** that can learn (Figure 1B).

SYNAPSE

The junction that connects two neurons. Its strength can vary with time depending on how often it is used to pass signal between the neurons.

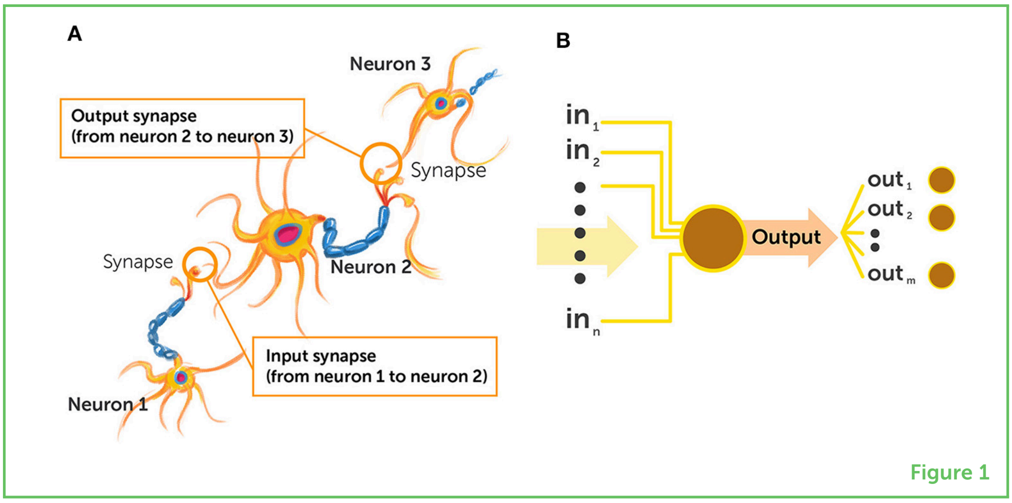

Figure 1

Biological and artificial neurons. **(A)** Biological neurons in the brain whose synapses' strength and location change when we learn. **(B)** Artificial neurons are digital units inside a computer, that receive input and transmit output to other artificial neurons.

ARTIFICIAL NEURONS

A representation of a neurons inside a computer. Artificial neurons receive digital inputs, perform a calculation, and transmit a digital output to other artificial neurons.

NEURAL NETWORKS—BIOLOGICAL AND ARTIFICIAL

A great leap in ML occurred with the birth of **neural networks (NNs)** [4]. A NN is a network of connected artificial neurons, simulated within a computer and organized in layers (to learn more about NNs, read this paper on artificial neural networks or this paper on the similarity between nerve cells and AI). The first layer, which is the input layer (Figure 2), receives the data that we feed it, for example, an image, video, or audio file. The middle layers, called the hidden layers, process the data, and the output layer gives the result of the computation (for example, identification of objects in an image or the transcription of voice to text). The NN is considered simple, or shallow, if it has only one hidden layer, and a deep neural network (DNN) if it has two or more hidden layers. In this article we will focus on DNN, which perform **deep learning** [5].

NEURAL NETWORK (NN)

A network of artificial computing elements that simulates the operation of neurons in the brain.

DEEP LEARNING (DL)

A machine learning technique based on neural networks with at least two internal layers.

Every artificial neuron in the DNN is connected to all other neurons in the next layer of the network, with various strengths of connection, called weights (Figure 3). Every neuron receives input from all other neurons in the previous layer, performs a calculation based on the weights of connections, and then sends an output to all other neurons in the next layer (Figure 3).

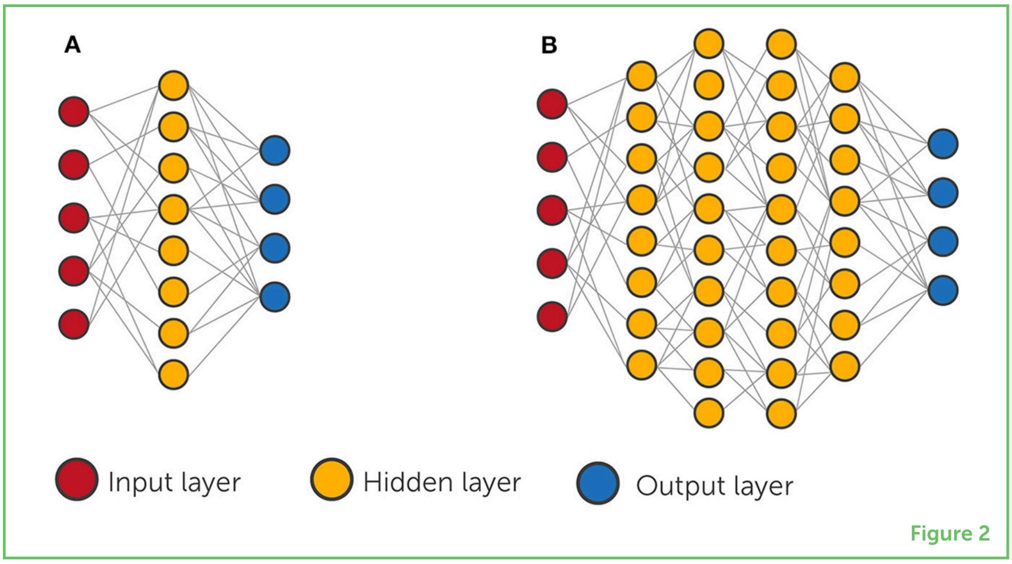

Figure 2

Artificial neural networks. Artificial neural networks are collections of artificial neurons organized in layers. **(A)** A simple neural network: one "hidden layer". **(B)** A deep neural network: two or more hidden layers. Large deep neural networks can contain about a thousand neurons in each layer and a few tens of hidden layers.

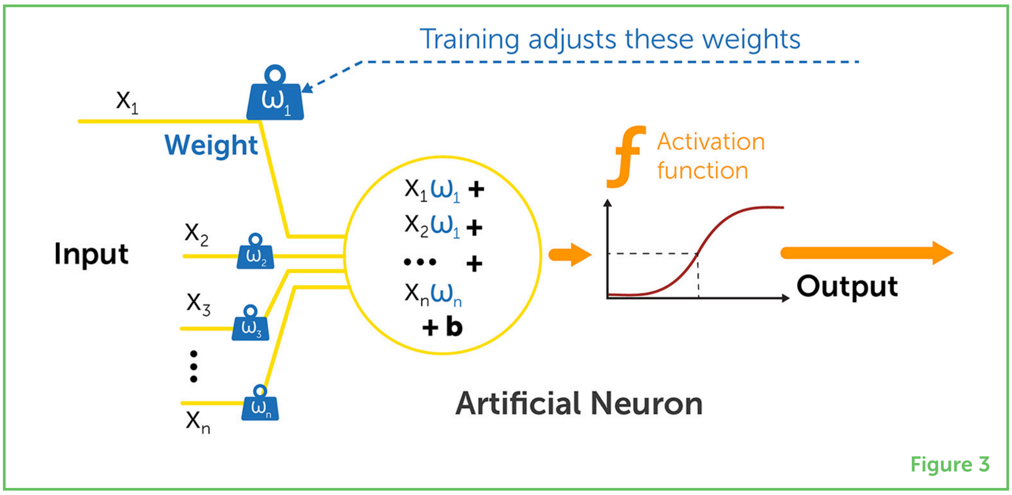

Figure 3

How neural networks learn. Every artificial neuron in a neural network receives input signals (x_1 to x_n) from the artificial neurons in the previous layer. It multiplies these inputs using a set of weights (w_1 to w_n), sums them up and adds another constant property called a bias (b), that determines how active a neuron will be. Finally, it runs this summation result through a function called an activation function, which determines the final result that the neuron sends, and sends the final output to all the neurons in the next layer. Learning occurs in the network by adjusting the set of weights and bias that the artificial neurons use in their calculations. To learn more about the calculations behind NNs, see here.

This way, information flows from the input layer to the output layer in what is called forward propagation. At the output layer, the DNN says what it thinks the right outcome is (for example, whether there is a cat or a dog in the picture).

Now comes an interesting part, where our network must learn. Learning means updating the weights in the DNN (like the changing synapses in the brain), to perform the task better next time. Think of every weight in the DNN as a knob which can be tuned to change the overall result—usually tens of thousands of them. To tune these knobs, the DNN first compares the outcome it got (the activation of its neurons in the output layer, each representing a specific outcome) with the right outcome it should have gotten (expected activation of each neuron in the output layer). Then, it updates the weights in all layers, from the last to the first hidden layer, to minimize the difference between the desired outcome and actual outcome.

There are two common ways to train DNNs. The first is supervised training, which is more accurate but less efficient, where systems scan many data samples already labeled by people (e.g., images of cats and dogs correctly categorized). The second way is self-supervised training, where the network tries to reconstruct information it already has (its input) after representing the information within its layers—this does not require human intervention.

BACKPROPAGATION

A learning algorithm commonly used for learning in deep neural networks.

The common algorithm used for updating DNN weights is called **backpropagation** [5, 6]. Backpropagation allows the DNN to improve its performance after every trial. The DNNs we currently train keep improving after every trial, but their rate of improvement decreases, so though at some point we stop investing in training, nonetheless the network keeps improving. After a network learns a training set of data, it then receives a test set to check its performance. When its performance is good enough for the specific task, it is ready to be used on new unseen data.

UNDERSTANDING THE WORLD USING CONVOLUTIONAL NEURAL NETWORKS

In the visual cortex—our brain's visual information processor—neurons respond to different visual features, called motifs, such as the orientation of lines [7], or edges. Collections of neurons, each detecting a different motif, are repeated everywhere inside the visual cortex, detecting the same motifs on different parts of the image we see. Inspired by this architecture of the brain, researchers applied the same basic structure on artificial neural nets [8]. I further developed these neural nets, which were trained using backpropagation [9] and that eventually led to the development of convolutional neural networks (CNNs).

The assumption behind CNNs is that signals we perceive from the world are composited of simple elements that build on top of each other to create complex objects. Simple motifs (e.g., lines and edges) are combined into parts of objects (e.g., square surfaces and legs), which are then combined into more complex objects (e.g., a table) and even categories of objects in a hierarchical manner (Figure 4). In a CNN, we take a set of neurons and connect them to a small patch in our input data, an image for example. Then, we take the same set of neurons and copy them all over the image, so that they look at every different part or patch. The neurons' output determines the presence or absence of a feature within each patch of the image. As the information propagates throughout the CNN layers, these features become more composite, or complex (to see an intuitive demonstration of how CNNs work, see this video). Amazingly, we do not have to tell the CNNs which features to look for—we just train them end-to-end using backpropagation, and they—quite magically—figure out themselves which features should be used to identify the object.

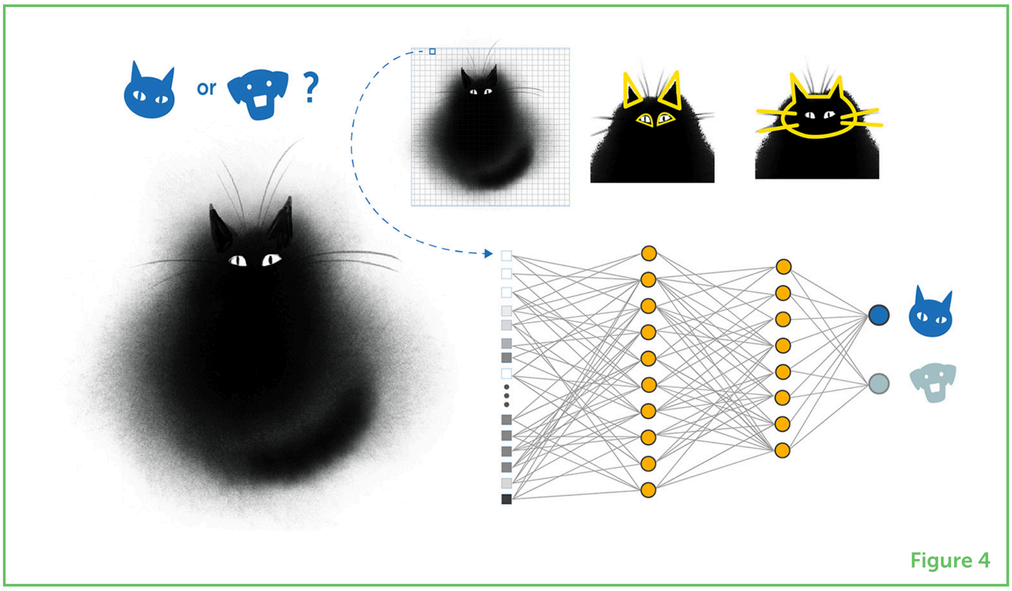

Figure 4

Figure 4

Image recognition using convolutional neural networks. Convolutional neural networks (CNNs) are used for various applications, including image recognition. The network represents an image in the input layer as a list of values (indicating the color of different pixels in the image). It then propagates the information into more complex objects (such as ears, eyes, whole face), until eventually the network output determines what object is present in the image (cat or dog in this example).

The first CNN I used was trained for recognizing hand-written digits [9]. It was quite successful and was later implemented for reading cheques in banks in the US and Europe. CNNs are useful for various other applications, related to images (such as Optical character recognition, or OCR, face recognition and video surveillance), as well as to speech recognition [10, 11]. CNNs have greatly improved the performance of former DNNs and are now integral to cutting-edge technologies such as medical imagery analysis and autonomous driving. My contribution to CNNs development won me the prestigious Turing Award in 2018, together with my colleagues Geoffrey Hinton and Yoshua Bengio.

CAN NEURAL NETWORKS HELP US UNDERSTAND THE BRAIN?

As we have seen, the development of artificial neural networks was inspired by the brain's operations. So can NNs be useful in understanding the brain itself? My answer is they are *necessary* to understand the brain. The brain is a very complex organ and there are probably some underlying principles governing its abilities that we have not uncovered yet. Though neuroscientists have collected a huge amount of experimental data, there is no solid theory about how the brain works. To show that a theory of brain functioning is correct, we need to recreate it within a computer and see that it works in ways that are somewhat similar to that of the brain. If we manage to build a computerized brain that acts similarly to the biological brain, this indicates we captured common operating principles, despite the differences between the systems.

Today, many brain scientists are already using deep learning and NNs as models for explaining brain activity [12], particularly in visual cortex, but it is also relevant for explaining how we process speech and text. NNs allow us to inspect information processes by recording the activity of all the neurons and using that to understand how they represent data. But to understand the collective operation of millions or billions of these elements, at some point we must describe the operation of the whole network at an abstract level. Having NNs that operate in similar ways to the brain would provide a great leap in our overall understanding of the brain, and of intelligence.

WILL MACHINES TAKE OVER THE WORLD?

I think the issue of controlling AI has become a modern "bogeyman", with horror forecasts of machines becoming smarter than us and dominating us (Figure 5). However, in humans and thus in machines, intelligence does not equal a will to dominate the other. Another concern is aligning behavior of intelligent machines with humanity's values. Although it is hard to "educate" machines to behave properly, we can manage it—the same way we educate our kids to behave in society, and with the same regulations to guide social functioning. We can define intrinsic objectives that machines will pursue (think of these like "core values") that they cannot violate or modify, ensuring that the machines' behavior stays aligned with our values and goals.

Figure 5

Figure 5

Should we be afraid of machines? I believe the worry about machine dominance is unfounded as we can ensure that machines will be friends not foes.

Every new technology brings with it some unexpected consequences, so we as a society must correct any unwanted side effects quickly, minimizing their damage. After online services developed, such as YouTube, Facebook, Instagram, we encountered the problem of improper content and developed means for content moderation. I am confident in our ability to deal with the problems of new technologies successfully as they arise.

What excites me the most about the future of AI is uncovering the underlying principles of intelligence. That would help us explain what human intelligence really is, and enable us to build intelligent systems, which would eventually expand human intelligence. Progressing in our understanding of the world requires more intelligence: at some point we will need other systems beyond our limited brains that we can use. For an interesting example of how AI helped us better understand the world, read this article about solving the long standing problem of protein folding.

Another, more engineering-based endeavor I want to see in the future is building intelligent systems to help us with our daily lives. For example, domestic robots that will be like intelligent human assistants, managing things we do not want to do and filtering out unimportant information. We call this an AI-complete problem [13], that requires the integration of many abilities and techniques. I work on new basic self-supervised learning algorithms that could hopefully bridge the gap between machine learning today and human learning. I hope we will be able to tackle AI-complete problems more successfully and live even more comfortable lives.

ADDITIONAL MATERIALS

1. Yann's Home Page.
2. Yann's Twitter Posts.
3. Check out this fake speech by the former president of the United States, Barack Obama, made using AI.
4. Can Computers Understand Humor?
5. DALL·E 2 and Lexica Aperture—AI-based online softwares for creating images from natural language.

ACKNOWLEDGMENTS

I wish to thank Or Raphael for conducting the interview which served as the basis for this paper, and for co-authoring the paper, and Alex Bernstein for providing the figures.

REFERENCES

1. Pfeifer, R., and Scheier, C. 2001. *Understanding Intelligence*. Cambridge: MIT Press.
2. Wolpert, D. H., and Macready, W. G. 1997. No free lunch theorems for optimization. *IEEE trans. Evolut. Computat*. 1:67–82. doi: 10.1109/4235.585893
3. Alpaydin, E. 2016. *Machine Learning: The New AI*. Cambridge: MIT press.

4. McCulloch, W. S., and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bullet. Mathemat. Biophys*. 5:115–33.
5. LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. *Nature*. 521:436–44. doi: 10.1038/nature14539
6. LeCun, Y., Touresky, D., Hinton, G., and Sejnowski, T. 1988. "A theoretical framework for back-propagation", in *Proceedings of the 1988 Connectionist Models Summer School* (Pittsburg, PA: Morgan Kaufmann), 21–28.
7. Hubel, D. H., and Wiesel, T. N. 1959. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148:574. doi: 10.1113/jphysiol.1959.sp006308
8. Fukushima, K., and Miyake, S. 1982. "Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition", in *Competition and Cooperation in Neural Nets* (Berlin: Springer), 267–285.
9. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., et al. 1989. "Handwritten digit recognition with a back-propagation network", in *Advances in Neural Information Processing Systems (NIPS 1989), Vol. 2* (Denver, CO: Morgan Kaufmann).
10. LeCun, Y., Kavukcuoglu, K., and Farabet, C. 2010. "Convolutional networks and applications in vision", in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (Paris: IEEE), 253–256.
11. LeCun, Y., and Bengio, Y. 1995. "Convolutional networks for images, speech, and time series", in *The Handbook of Brain Theory and Neural Networks*, ed. M. A. Arbib (MIT Press).
12. Yamins, D. L., and DiCarlo, J. J. 2016. Using goal-driven deep learning models to understand sensory cortex. *Nature Neurosci.* 19:356–365. Available online at: https://www.nature.com/articles/nn.4244
13. Weston, J., Bordes, A., Chopra, S., Rush, A. M., Van Merriënboer, B., Joulin, A., et al. 2015. "Towards AI-complete question answering: a set of prerequisite toy tasks", in *arXiv*.

SUBMITTED: 04 April 2023; **ACCEPTED:** 26 March 2024;
PUBLISHED ONLINE: 07 May 2024.

EDITOR: Idan Segev, Hebrew University of Jerusalem, Israel

SCIENCE MENTORS: Yunchao Tang and Tomas Emmanuel Ward

CITATION: LeCun Y (2024) Will Learning Machines Take Over the World? Front. Young Minds 12:1164958. doi: 10.3389/frym.2024.1164958

CONFLICT OF INTEREST: YL was employed by Meta.

YOUNG REVIEWERS

OISIN, AGE: 12

Oisin likes to play piano, chess, and play video games with his friends. He also like to draw and play soccer. His favorite video games are Minecraft, City Skylines, and Civilization VI. Oisin loves to read Harry Potter and Discworld books by Terry Pratchett. Oisin lives in Ireland is in sixth class in national school.

ZI-AN, AGE: 8

Hi, I am Zi-An, coming from a family of teachers, I have inherited a love for knowledge and learning. But my biggest joy? That is definitely my little brother. I absolutely love goofing, around and making him laugh! I am fascinated about science. I want to research the secrets of everlasting life, so that people I love will never grow old or die.

AUTHORS

YANN LECUN

Yann LeCun is VP & Chief AI Scientist at Meta and Silver Professor at NYU affiliated with the Courant Institute of Mathematical Sciences & the Center for Data Science. Yann received an Engineering Diploma from ESIEE (Paris) and a PhD in Computer Science from Sorbonne Université. After a postdoc in Toronto, Yann joined AT&T Bell Labs in 1988, and in 1996 he became Head of the Image Processing Lab. at AT&T. He joined NYU as a professor in 2003 and Meta/Facebook in 2013. His interests include AI machine learning, computer perception, robotics and computational neuroscience. Yann won several prestigious awards including the Turing award (2018), AAAI Fellow (2019), and Legion of Honor (2020), and is a member of the National Academy of Sciences, the National Academy of Engineering, the French Académie des Sciences. Yann has several hobbies, including building model airplanes with his family and playing music (mostly Jazz nowadays). *yann@cs.nyu.edu