CHAPTER

4

Deep Learning

The Revolution in AI

Take any old classification problem where you have a lot of data, and it's going to be solved by deep learning. There's going to be thousands of applications of deep learning.

—Geoffrey Hinton,
English Canadian cognitive psychologist and computer scientist[1]

Fei-Fei Li, who got a BA degree in physics from Princeton in 1999 with high honors and a PhD in electrical engineering from Caltech in 2005, focused her brilliance on developing AI models. But she had a major challenge: finding quality datasets. At first, she looked at creating them by hand, such as with graduate students who downloaded images from the Internet. But the process was too slow and tedious.

One day a student mentioned to Li that Amazon.com's Mechanical Turk, an online service that uses crowdsourcing to solve problems, could be a good way to scale the process. It would allow for fast and accurate labeling of the data.

Li gave it a try, and it worked out quite well. By 2010, she had created ImageNet, which had 3.2 million images across over 5,200 categories.

[1] Siddhartha Mukherjee, "The Algorithm Will See You Now," *The New Yorker*, April 3, 2017, https://www.newyorker.com/magazine/2017/04/03/ai-versus-md.

Yet it got a tepid response from the academic community. But this did not deter Li. She continued to work tirelessly to evangelize the dataset. In 2012, she put together a contest as a way to encourage researchers to create more effective models and push the boundaries of innovation. It would turn out to be a game changer, and the contest would become an annual event.

In the first contest, professors from the University of Toronto—Geoffrey Hinton, Ilya Sutskever, and Alex Krizhevsky—used sophisticated deep learning algorithms. And the results were standout. The system they built, which was called AlexNet, beat all the other contestants by a margin of 10.8%.[2]

This was no fluke. In the years after this, deep learning continued to show accelerated progress with ImageNet. As of now, the error rate for deep learning is a mere 2% or so—which is better than humans.

By the way, Li has since gone on to become a professor at Stanford and co-director of the school's AI lab. She is also Google's chief scientist of AI and Machine Learning. Needless to say, whenever she has new ideas now, people listen!

In this chapter, we'll take a look at deep learning, which is clearly the hottest area of AI. It has led to major advances in areas like self-driving cars and virtual assistants like Siri.

Yes, deep learning can be a complicated subject, and the field is constantly changing. But we'll take a look at the main concepts and trends—without getting into the technical details.

Difference Between Deep Learning and Machine Learning

There is often confusion between deep learning and machine learning. And this is reasonable. Both topics are quite complex, and they do share many similarities.

So to understand the differences, let's first take a look at two high-level aspects of machine learning and how they relate to deep learning. First of all, while both usually require large amounts of data, the types are generally different.

Take the following example: Suppose we have photos of thousands of animals and want to create an algorithm to find the horses. Well, machine learning cannot analyze the photos themselves; instead, the data must be labeled. The machine learning algorithm will then be trained to recognize horses, through a process known as supervised learning (covered in Chapter 3).

Even though machine learning will likely come up with good results, they will still have limitations. Wouldn't it be better to look at the pixels of the images themselves—and find the patterns? Definitely.

But to do this with machine learning, you need to use a process called feature extraction. This means you must come up with the kinds of characteristics of a horse—such as the shape, the hooves, color, and height—which the algorithms will then try to identify.

Again, this is a good approach—but it is far from perfect. What if your features are off the mark or do not account for outliers or exceptions? In such cases, the accuracy of the model will likely suffer. After all, there are many variations to a horse. Feature extraction also has the drawback of ignoring a large amount of the data. This can be exceedingly complicated—if not impossible—for certain use cases. Look at computer viruses. Their structures and patterns, which are known as signatures, are constantly changing so as to infiltrate systems. But with feature extraction, a person would somehow have to anticipate this, which is not practical. This is why cybersecurity software is often about collecting signatures after a virus has exacted damage.

But with deep learning, we can solve these problems. This approach analyzes all the data—pixel by pixel—and then finds the relationships by using a neural network, which mimics the human brain.

Let's take a look.

So What Is Deep Learning Then?

Deep learning is a subfield of machine learning. This type of system allows for processing huge amounts of data to find relationships and patterns that humans are often unable to detect. The word "deep" refers to the number of hidden layers in the neural network, which provide much of the power to learn.

When it comes to the topic of AI, deep learning is at the cutting-edge and often generates most of the buzz in mainstream media. "[Deep learning] AI is the new electricity," extolled Andrew Yan-Tak Ng, who is the former chief scientist at Baidu and co-founder of Google Brain.[3]

But it is also important to remember that deep learning is still in the early stages of development and commercialization. For example, it was not until about 2015 that Google started using this technology for its search engine.

As we saw in Chapter 1, the history of neural networks was full of ebbs and flows. It was Frank Rosenblatt who created the perceptron, which was a fairly basic system. But real academic progress with neural networks did not occur until the 1980s, such as with the breakthroughs with backpropagation, convolutional neural networks, and recurrent neural networks. But for deep learning to have an impact on the real world, it would take the staggering growth in data, such as from the Internet, and the surge in computing power.

The Brain and Deep Learning

Weighing only about 3.3 pounds, the human brain is an amazing feat of evolution. There are about 86 billion neurons—often called gray matter—that are connected with trillions of synapses. Think of neurons as CPUs (Central Processing Units) that take in data. The learning occurs with the strengthening or weakening of the synapses.

The brain is made up of three regions: the forebrain, the midbrain, and the hindbrain. Among these, there are a variety of areas that perform different functions. Some of the main ones include the following:

- *Hippocampus*: This is where your brain stores memories. In fact, this is the part that fails when a person has Alzheimer's disease, in which a person loses the ability to form short-term memories.
- *Frontal Lobe*: Here the brain focuses on emotions, speech, creativity, judgment, planning, and reasoning.
- *Cerebral Cortex*: This is perhaps the most important when it comes to AI. The cerebral cortex helps with thinking and other cognitive activities. According to research from Suzana Herculano-Houzel, the level of intelligence is related to the number of neurons in this area of the brain.

Then how does deep learning compare to the human brain? There are some tenuous similarities. At least in areas like the retina, there is a process of ingesting data and processing them through a complex network, which is based on assigning weights. But of course, this is only a minute part of the learning process. Besides, there are still many mysteries about the human brain, and of course, it is not based on things like digital computing (instead, it appears that it is more of an analogue system). However, as the research continues to advance, the discoveries in neuroscience could help build new models for AI.

Artificial Neural Networks (ANNs)

At the most basic level, an artificial neural network (ANN) is a function that includes units (which may also be called neurons, perceptrons, or nodes). Each unit will have a value and a weight, which indicates the relative importance, and will go into the hidden layer. The hidden layer uses a function, with the result becoming the output. There is also another value, called bias, which is a constant and is used in the calculation of the function.

This type of training of a model is called a feed-forward neural network. In other words, it only goes from input to the hidden layer to the output. It does not cycle back. But it could go to a new neural network, with the output becoming the input.

Figure 4-1 shows a chart of a feed-forward neural network.

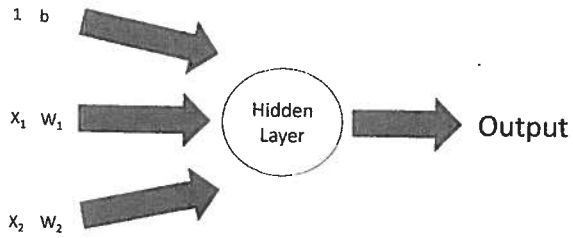

Figure 4-1. A basic feed-forward neural network

Let's go deeper on this by taking an example. Suppose you are creating a model to predict whether a company's stock will increase. The following are what the variables represent as well as the values and weights assigned:

- X_1: Revenues are growing at a minimum of 20% a year. The value is 2.
- X_2: The profit margin is at least 20%. The value is 4.
- W_1: 1.9.
- W_2: 9.6.
- b: This is the bias (the value is 1), which helps smooth out the calculations.

You'll then sum the weights, and then the function will process the information. This will often involve an activation function, which is non-linear. This is more reflective of the real world since data is usually not in a straight line.

Now there are a variety of activation functions to choose from. One of the most common is the sigmoid. This compresses the input value into a range of 0–1. The closer it is to 1, the more accurate the model.

When you graph this function, it will look like an S shape. See Figure 4-2.

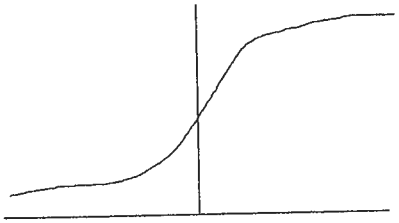

Figure 4-2. A typical sigmoid activation function

As you can see, the system is relatively simplistic and will not be helpful in high-end AI models. To add much more power, there usually needs to be multiple hidden layers. This results in a multilayered perceptron (MLP). It also helps to use something called backpropagation, which allows for the output to be circled back into the neural network.

Backpropagation

One of the major drawbacks with artificial neural networks is the process of making adjustments to the weights in the model. Traditional approaches, like the use of the mutation algorithm, used random values that proved to be time consuming.

Given this, researchers looked for alternatives, such as backpropagation. This technique had been around since the 1970s but got little interest as the performance was lacking. But David Rumelhart, Geoffrey Hinton, and Ronald Williams realized that backpropagation still had potential, so long as it was refined. In 1986, they wrote a paper entitled "Learning Representations by Back-propagating Errors," and it was a bombshell in the AI community.[4] It clearly showed that backpropagation could be much faster but also allow for more powerful artificial neural networks.

As should be no surprise, there is a lot of math involved in backpropagation. But when you boil things down, it's about adjusting the neural network when errors are found and then iterating the new values through the neural network again. Essentially, the process involves slight changes that continue to optimize the model.

For example, let's say one of the inputs has an output of 0.6. This means that the error is 0.4 (1.0 minus 0.6), which is subpar. But we can then backpropogate the output, and perhaps the new output may get to 0.65. This training will go on until the value is much closer to 1.

Figure 4-3 illustrates this process. At first, there is a high level of errors because the weights are too large. But by making iterations, the errors will gradually fall. However, doing too much of this could mean an increase in errors. In other words, the goal of backpropagation is to find the midpoint.

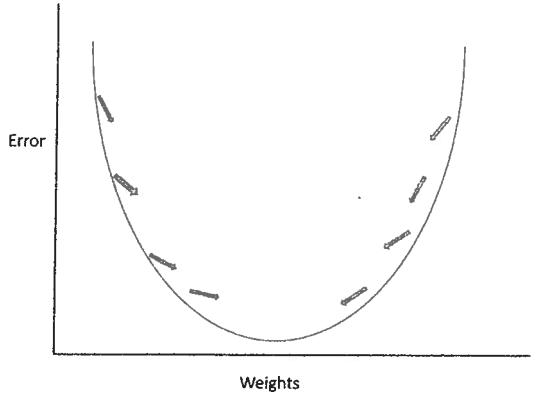

Figure 4-3. The optimal value for a backpropagation function is at the bottom of the graph

As a gauge of the success of backpropagation, there were a myriad of commercial applications that sprung up. One was called NETtalk, which was developed by Terrence Sejnowski and Charles Rosenberg in the mid-1980s. The machine was able to learn how to pronounce English text. NETtalk was so interesting that it was even demoed on the *Today show*.

There were also a variety of startups that were created that leveraged backpropagation, such as HNC Software. It built models that detected credit card fraud. Up until this point—when HNC was founded in the late 1980s—the process was done mostly by hand, which led to costly errors and low volumes of issuances. But by using deep learning approaches, credit card companies were able to save billions of dollars.

In 2002, HNC was acquired by Fair, Isaac and valued at $810 million.[5]

The Various Neural Networks

The most basic type of a neural network is a fully connected neural network. As the name implies, it is where all the neurons have connections from layer to layer. This network is actually quite popular since it means having to use little judgment when creating the model.

Then what are some of the other neural networks? The common ones include the recurrent neural network (RNN), the convolutional neural network (CNN), and the generative adversarial network (GAN), which we'll cover next.

Recurrent Neural Network

With a recurrent neural network (RNN), the function not only processes the input but also prior inputs across time. An example of this is what happens when you enter characters in a messaging app. As you begin to type, the system will predict the words. So if you tap out "He," the computer will suggest "He," "Hello," and "Here's." The RNN is essentially a string of neural networks that feed on each other based on complex algorithms.

There are variations on the model. One is called LSTM, which stands for long short-term memory. This came about from a paper written by professors Sepp Hochreiter and Jürgen Schmidhuber in 1997.[6] In it, they set forth a way to effectively use inputs that are separated from each other for long time periods, allowing the use of more datasets.

Of course, RNNs do have drawbacks. There is the vanishing gradient problem, which means that the accuracy decays as the models get larger. The models can also take longer to train.

To deal with this, Google developed a new model called the Transformer, which is much more efficient since it processes the inputs in parallel. It also results in more accurate results.

Google has gained much insight about RNNs through its Translate app, which handles over 100 languages and processes over 100 billion words a day.[7] Launched in 2006, it initially used machine learning systems. But in 2016, Google switched to deep learning by creating Google Neural Machine Translation.[8] All in all, it has resulted in much higher accuracy rates.[9]

[6] Sepp Hochreiter and Jürgen Schmidhuber, "Long Short-Term Memory," *Neural Computation* 9, no. 8 (1997): 1735-80.
[7] www.argotrans.com/blog/accurate-google-translate-2018/
[8] www.techspot.com/news/75637-google-translate-not-monetized-despite-

Consider how Google Translate has helped out doctors who work with patients who speak other languages. According to a study from the University of California, San Francisco (UCSF), that was published in *JAMA Internal Medicine*, the app had a 92% accuracy rate with English-to-Spanish translations. This was up from 60% over the past couple years.[10]

Convolutional Neural Network (CNN)

Intuitively, it makes sense to have all the units in a neural network to be connected. This works well with many applications.

But there are scenarios where it is far from optimal, such as with image recognition. Just imagine how complex a model would be where every pixel is a unit! It could quickly become unmanageable. There would also be other complications like overfitting. This is where the data is not reflective of what is being tested or there is a focus on the wrong features.

To deal with all this, you can use a convolutional neural network (CNN). The origins of this go back to professor Yann LeCun in 1998, when he published a paper called "Gradient-Based Learning Applied to Document Recognition."[11] Despite its strong insights and breakthroughs, it got little traction. But as deep learning started to show significant progress in 2012, researchers revisited the model.

LeCun got his inspiration for the CNN from Nobel Prize winners David Hubel and Torsten Wiesel who studied neurons of the visual cortex. This system takes an image from the retina and processes it in different stages—from easy to more complex. Each of the stages is called a convolution. For example, the first level would be to identify lines and angles; next, the visual cortex will find the shapes; and then it will detect the objects.

This is analogous to how a computer-based CNN works. Let's take an example: Suppose you want to build a model that can identify a letter. The CNN will have input in the form of an image that has 3,072 pixels. Each of the pixels will have a value that is from 0 to 255, which indicates the overall intensity. By using a CNN, the computer will go through multiple variations to identify the features.

The first is the convolutional layer, which is a filter that scans the image. In our example, this could be 5 × 5 pixels. The process will create a feature map, which is a long array of numbers. Next, the model will apply more filters to the image. By doing this, the CNN will identify the lines, edges and shapes—all

[10] https://gizmodo.com/google-translate-can-help-doctors-bridge-the-language-g-1832881294

expressed in numbers. With the various output layers, the model will use pooling, which combines them to generate a single output, and then create a fully connected neural network.

A CNN can definitely get complex. But it should be able to accurately identify the numbers that are input into the system.

Generative Adversarial Networks (GANs)

Ian Goodfellow, who got his masters in computer science at Stanford and his PhD in machine learning at the Université de Montréal, would go on to work at Google. In his 20s, he co-authored one of the top books in AI, called *Deep Learning*,[12] and also made innovations with Google Maps.

But it was in 2014 that he had his most impactful breakthrough. It actually happened in a pub in Montreal when he talked with some of his friends about how deep learning could *create* photos.[13] At the time, the approach was to use generative models, but they were often blurry and nonsensical.

Goodfellow realized that there had to be a better why. So why not use game theory? That is, have two models compete against each other in a tight feedback loop. This could also be done with unlabeled data.

Here's a basic workflow:

- *Generator*: This neural network creates a myriad of new creations, such as photos or sentences.
- *Discriminator*: This neural network looks at the creations to see which ones are real.
- *Adjustments*: With the two results, a new model would change the creations to make them as realistic as possible. Through many iterations, the discriminator will no longer need to be used.

He was so excited about the idea that after he left the pub he started to code his ideas. The result was a new deep learning model: the generative adversarial network or GAN. And the results were standout. He would soon become an AI rock star.

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (Cambridge, MA: The MIT Press, 2016).

GAN research has already spurred over 500 academic papers.[14] Companies like Facebook have also used this technology, such as for its photo analysis and processing. The company's chief AI scientist, Yann LeCun, noted that GANs are the "the coolest idea in deep learning in the last 20 years."[15]

GANs have also been shown to help with sophisticated scientific research. For example, they have helped improve the accuracy of detecting behavior of subatomic particles in the Large Hadron Collider at CERN in Switzerland.[16]

While still in the early innings, this technology could lead to such things as a computer that can develop new types of fashion items or maybe a new-fangled wearable. Perhaps a GAN could even come up with a hit rap song.

And it could be sooner than you think. As a teenager, Robbie Barrat taught himself how to use deep learning systems and built a model to rap in the style of Kanye West.

But this was just the beginning of his AI wizardry. As a researcher at Stanford, he developed his own GAN platform, which processed roughly 10,000 nude portraits. The system then would create truly mesmerizing new works of art (you can find them at his Twitter account at @DrBeef_).

Oh, and he also made his system open source at his GitHub account. This caught the attention of a collective of French artists, called Obvious, that used the technology to create portraits of an eighteenth-century fictional family. It was based on processing 15,000 portraits from the fourteenth to the twentieth centuries.

In 2018, Obvious put its artwork at a Christie's auction, fetching a cool $432,000. [17]

But unfortunately, when it comes to GANs, there have been uses that have been less than admirable. One example is to use them for deepfakes, which involve leveraging neural networks to create images or videos that are misleading. Some of this is just kind of playful. For example, one GAN makes it possible to have Barack Obama say anything you tell him!

Yet there are lots of risks. Researchers at New York University and the Michigan State University wrote a paper that focused on "DeepMasterPrints."[18]

[4] https://github.com/hindupuravinash/the-gan-zoo

[5] https://trendsandevents4developers.wordpress.com/2017/04/24/the-coolest-idea-in-deep-learning-in-20-years-and-more/

[6] www.hpcwire.com/2018/08/14/cern-incorporates-ai-into-physics-based-simulations/

[7] www.washingtonpost.com/nation/2018/10/26/year-old-developed-code-ai-portrait-that-sold-christies/?utm_term=.b2f366a4460e

It showed how a GAN can develop fake fingerprints to unlock three types of smartphones!

Then there was the incident of a so-called deepfake video of actress Jennifer Lawrence at a Golden Globes press conference. Her face was merged with Steve Buscemi's.[19]

Deep Learning Applications

With so much money and resources being devoted to deep learning, there has been a surge in innovations. It seems that every day there is something amazing that is being announced.

Then what are some of the applications? Where has deep learning proven to be a game changer? Let's take a look at some that cover areas like healthcare, energy, and even earthquakes

Use Case: Detecting Alzheimer's Disease

Despite decades of research, a cure for Alzheimer's disease remains elusive. Although, scientists have developed drugs that have slowed down the progression of the disease.

In light of this, early diagnosis is critical—and deep learning can potentially be a big help. Researchers at the UCSF Department of Radiology and Biomedical Imaging have used this technology to analyze brain screens—from the Alzheimer's Disease Neuroimaging Initiative public dataset—and to detect changes in the levels of glucose.

The result: The model can diagnose Alzheimer's disease up to six years before a clinical diagnosis. One of the tests showed a 92% accuracy rate, and another was 98%.

Now this is still in the beginning phases—and there will need to be more datasets analyzed. But so far, the results are very encouraging.

According to Dr. Jae Ho Sohn, who authored the study:

> This is an ideal application of deep learning because it is particularly strong at finding very subtle but diffuse processes. Human radiologists are really strong at identifying tiny focal finding like a brain tumor, but we struggle at detecting more slow, global changes. Given the strength of deep learning in this type of application, especially compared to humans, it seemed like a natural application.[20]

Use Case: Energy

Because of its massive data center infrastructure, Google is one of the largest consumers of energy. Even a small improvement in efficiency can lead to a sizeable impact on the bottom line. But there could also be the benefits of less carbon emissions.

To help with these goals, Google's DeepMind unit has been applying deep learning, which has involved better management of wind power. Even though this is a clean source of energy, it can be tough to use because of the changes in weather.

But DeepMind's deep learning algorithms have been critical. Applied to 700 megawatts of wind power in the United States, they were able to make accurate forecasts for output with a lead time of 36 hours.

According to DeepMind's blog:

> This is important, because energy sources that can be scheduled (i.e. can deliver a set amount of electricity at a set time) are often more valuable to the grid...To date, machine learning has boosted the value of our wind energy by roughly 20 percent, compared to the baseline scenario of no time-based commitments to the grid.[21]

But of course, this deep learning system could be more than just about Google—it could have a wide-ranging impact on energy use across the world.

Use Case: Earthquakes

Earthquakes are extremely complicated to understand. They are also exceedingly difficult to predict. You need to evaluate faults, rock formations and deformations, electromagnetic activity, and changes in the groundwater. Hey, there is even evidence that animals have the ability to sense an earthquake!

But over the decades, scientists have collected huge amounts of data on this topic. In other words, this could be an application for deep learning, right?

Absolutely.

Seismologists at Caltech, which include Yisong Yue, Egill Hauksson, Zachary Ross, and Men-Andrin Meier, have been doing considerable research on this, using convolutional neural networks and recurrent neural networks. They are trying to build an effective early-warning system.

Here's what Yue had to say:

> AI can [analyze earthquakes] faster and more accurately than humans can, and even find patterns that would otherwise escape the human eye. Furthermore, the patterns we hope to extract are hard for rule-based systems to adequately capture, and so the advanced pattern-matching abilities of modern deep learning can offer superior performance than existing automated earthquake monitoring algorithms.[22]

But the key is improving data collection. This means more analysis of small earthquakes (in California, there is an average of 50 each day). The goal is to create an earthquake catalog that can lead to the creation of a virtual seismologist, who can make evaluations of an earthquake faster than a human. This could allow for faster lead times when an earthquake strikes, which may help to save lives and property.

Use Case: Radiology

PET scans and MRIs are amazing technology. But there are definitely downsides. A patient needs to stay within a confining tube for 30 minutes to an hour. This is uncomfortable and means being exposed to gadolinium, which has been shown to have harmful side effects.

Greg Zaharchuk and Enhao Gong, who met at Stanford, thought there could be a better way. Zaharchuk was an MD and PhD, with a specialization in radiology. He was also the doctoral advisor of Gong, who was an electrical engineering PhD in deep learning and medical image reconstruction.

In 2017, they co-founded Subtle Medical and hired some of the brightest imaging scientists, radiologists, and AI experts. Together, they set themselves to the challenge of improving PET scans and MRIs. Subtle Medical created a system that not only reduces the time for an MRI and PET scans by up to ten times, but the accuracy has been much higher. This was powered by high-end NVIDIA GPUs.

Then in December 2018, the system received FDA (Federal Drug Administration) 510(k) clearance and a CE mark approval for the European market.[23] It was the first ever AI-based nuclear medical device to achieve both of these designations.

[22] www.caltech.edu/about/news/qa-creating-virtual-seismologist-84789

Subtle Medical has more plans to revolutionize the radiology business. As of 2019, it is developing SubtleMRTM, which will be even more powerful than the company's current solution, and SubtleGADTM, which will reduce gadolinium dosages.[24]

Deep Learning Hardware

Regarding chip systems for deep learning, GPUs have been the primary choice. But as AI gets more sophisticated—such as with GANs—and the datasets much larger, there is certainly more room for new approaches. Companies also have custom needs, such as in terms of functions and data. After all, an app for a consumer is usually quite different than one that is focused on the enterprise.

As a result, some of the mega tech companies have been developing their own chipsets:

- *Google*: In the summer of 2018, the company announced its third version of its Tensor Processing Unit (TPU; the first chip was developed in 2016).[25] The chips are so powerful—handling over 100 petaflops for training of models—there needs to be liquid cooling in the data centers. Google has also announced a version of its TPU for devices. Essentially, it means that processing will have less latency because there will be no need to access the cloud.
- *Amazon*: In 2018, the company announced AWS Inferentia.[26] The technology, which has come out of the acquisition of Annapurna in 2015, is focused on handling complex inference operations. In other words, this is what happens after a model has been trained.
- *Facebook and Intel*: These companies have joined forces to create an AI chip.[27] But the initiative is still in the initial phases. Intel also has been getting traction with an AI chip called the Nervana Neural Network Processor (NNP).

[24] www.streetinsider.com/Press+Releases/Subtle+Medical+Receives+FDA+510%28k%29+Clearance+and+CE+Mark+Approval+for+SubtlePET™/14892974.html
[25] www.theregister.co.uk/2018/05/09/google_tpu_3/
[26] https://aws.amazon.com/about-aws/whats-new/2018/11/announcing-amazon-inferentia-machine-learning-inference-microchip/

- *Alibaba*: The company has created its own AI chip company called Pingtouge.[28] It also has plans to build a quantum computer processor, which is based on qubits (they represent subatomic particles like electrons and photons).
- *Tesla*: Elon Musk has developed his own AI chip. It has 6 billion transistors and can process 36 trillion operations per second.[29]

There are a variety of startups that are making a play for the AI chip market as well. Among the leading companies is Untether AI, which is focused on creating chips that boost the transfer speeds of data (this has been a particularly difficult part of AI). In one of the company's prototypes, this process was more than 1,000 faster than a typical AI chip.[30] Intel, along with other investors, participated in a $13 million round of funding in 2019.

Now when it comes to AI chips, NVIDIA has the dominant market share. But because of the importance of this technology, it seems inevitable that there will be more and more offerings that will come to the market.

When to Use Deep Learning?

Because of the power of deep learning, there is the temptation to use this technology first when creating an AI project. But this can be a big mistake. Deep learning still has narrow use cases, such as for text, video, image, and time-series datasets. There is also a need for large amounts of data and high-powered computer systems.

Oh, and deep learning is better when outcomes can be quantified and verified.

To see why, let's consider the following example. A team of researchers, led by Thomas Hartung (a toxicologist at Johns Hopkins University), created a dataset of about 10,000 chemicals that were based on 800,000 animal tests. By using deep learning, the results showed that the model was more predictive than many animal tests for toxicity.[31] Remember that animal tests can not only be costly and require safety measures but also have inconsistent results because of repeated testing on the same chemical.

[28] www.technologyreview.com/s/612190/why-alibaba-is-investing-in-ai-chips-and-quantum-computing/
[29] www.technologyreview.com/f/613403/tesla-says-its-new-self-driving-chip-will-help-make-its-cars-autonomous/
[30] www.technologyreview.com/f/613258/intel-buys-into-an-ai-chip-

"The first scenario illustrates the predictive power of deep learning, and its ability to unearth correlations from large datasets that a human would never find," said Sheldon Fernandez, who is the CEO of DarwinAI.[32]

So where's a scenario in which deep learning falls short? Actually, an illustration of this is the 2018 FIFA World Cup in Russia, which France won. Many researchers tried to predict the outcomes of all 64 matches, but the results were far from accurate:[33]

- One group of researchers employed the bookmaker consensus model that indicated that Brazil would win.
- Another group of researchers used algorithms such as random forest and Poisson ranking to forecast that Spain would prevail.

The problem here is that it is tough to find the right variables that have predictive power. In fact, deep learning models are basically unable to handle the complexity of features for certain events, especially those that have elements of being chaotic.

However, even if you have the right amount of data and computing power, you still need to hire people who have a background in deep learning, which is not easy. Keep in mind that it is a challenge to select the right model and fine-tune it. How many hyperparameters should there be? What should be the number of hidden layers? And how do you evaluate the model? All of these are highly complex.

Even experts can get things wrong. Here's the following from Sheldon:

> One of our automotive clients encountered some bizarre behavior in which a self-driving car would turn left with increasing regularity when the sky was a certain shade of purple. After months of painful debugging, they determined the training for certain turning scenarios had been conducted in the Nevada desert when the sky was a particular hue. Unbeknownst to its human designers, the neural network had established a correlation between its turning behavior and the celestial tint.[34]

There are some tools that are helping with the deep learning process, such as Amazon.com's SageMaker, Google's HyperTune, and SigOpt. But there is still a long way to go.

If deep learning is not a fit, then you may want to consider machine learning, which often requires relatively less data. Furthermore, the models tend to be much simpler, but the results may still be more effective.

[32] This is from the author's interview with Sheldon Fernandez, the CEO of DarwinAI.
[33] https://medium.com/futuristone/artificial-intelligence-failed-

Drawbacks with Deep Learning

Given all the innovations and breakthroughs, it's reasonable that many people consider deep learning to be a silver bullet. It will mean we no longer have to drive a car. It may even mean that we'll cure cancer.

How is it not possible to be excited and optimistic? This is natural and reasonable. But it is important to note that deep learning is still in a nascent stage and there are actually many nagging issues. It's a good idea to temper expectations.

In 2018, Gary Marcus wrote a paper entitled "Deep Learning: A Critical Appraisal," in which he clearly set forth the challenges.[35] In his paper, he notes:

> Against a background of considerable progress in areas such as speech recognition, image recognition, and game playing, and considerable enthusiasm in the popular press, I present ten concerns for deep learning, and suggest that deep learning must be supplemented by other techniques if we are to reach Artificial General Intelligence.[36]

Marcus definitely has the right pedigree to present his concerns, as he has both an academic and business background in AI. Before becoming a professor at the Department of Psychology at New York University, he sold his startup, called Geometric Intelligence, to Uber. Marcus is also the author of several bestselling books like *The Haphazard Construction of the Human Mind*.[37]

Here's a look at some of his worries about deep learning:

- *Black Box*: A deep learning model could easily have millions of parameters that involve many hidden layers. Having a clear understanding of this is really beyond a person's capabilities. True, this may not necessarily be a problem with recognizing cats in a dataset. But it could definitely be an issue with models for medical diagnosis or determining the safety of an oil rig. In these situations, regulators will want to have a good understanding of the transparency of the models. Because of this, researchers are looking at creating systems to determine "explainability," which provides an understanding of deep learning models.

[35] Gary Marcus, "Deep Learning: A Critical Appraisal," arXiv, 1801.00631v1 [cs.AI]:1–27, 2018.
[36] https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf

- *Data*: The human brain has its flaws. But there are some functions that it does extremely well like the ability to learn by abstraction. For example, suppose Jan, who is five years old, goes to a restaurant with her family. Her mother points out an item on the plate and says it is a "taco." She does not have to explain it or provide any information about it. Instead, Jan's brain will instantly process this information and understand the overall pattern. In the future, when she sees another taco—even if it has differences, such as with the dressing—she will know what it is. For the most part, this is intuitive. But unfortunately, when it comes to deep learning, there is no taco learning by abstraction! The system has to process enormous amounts of information to recognize it. Of course, this is not a problem for companies like Facebook, Google, or even Uber. But many companies have much more limited datasets. The result is that deep learning may not be a good option.
- *Hierarchical Structure*: This way of organizing does not exist in deep learning. Because of this, language understanding still has a long way to go (especially with long discussions).
- *Open-Ended Inference*: Marcus notes that deep learning cannot understand the nuances between "John promised Mary to leave" and "John promised to leave Mary." What's more, deep learning is far away from being able to, for instance, read Jane Austen's *Pride and Prejudice* and be able to divine Elizabeth Bennet's character motivations.
- *Conceptual Thinking*: Deep learning cannot have an understanding of concepts like democracy, justice, or happiness. It also does not have imagination, thinking of new ideas or plans.
- *Common Sense*: This is something deep learning does not do well. If anything, this means a model can be easily confused. For example, let's say you ask an AI system, "Is it possible to make a computer with a sponge?" For the most part, it will probably not know that this is a ridiculous question.
- *Causation*: Deep learning is unable to determine this. It's all about finding correlations.

- *Prior Knowledge*: CNNs can help with some prior information, but this is limited. Deep learning is still fairly self-contained, as it only solves one problem at a time. It cannot take in the data and create algorithms that span various domains. In addition, a model does not adapt. If there is change in the data, then a new model needs to be trained and tested. And finally, deep learning does not have prior understanding of what people know instinctively—such as basic physics of the real world. This is something that has to be explicitly programmed into an AI system.
- *Static*: Deep learning works best in environments that are fairly simple. This is why AI has been so effective with board games, which have a clear set of rules and boundaries. But the real world is chaotic and unpredictable. This means that deep learning may fall short with complex problems, even with self-driving cars.
- *Resources*: A deep learning model often requires a tremendous amount of CPU power, such as with GPUs. This can get costly. Although, one option is to use a third-party cloud service.

This is quite a lot? It's true. But the paper still has left out some drawbacks. Here are a couple other ones:

- *Butterfly Effect*: Because of the complexity of the data, networks, and connections, a minute change can have a major impact in the results of the deep learning model. This could easily lead to conclusions that are wrong or misleading.
- *Overfitting*: We explained this concept earlier in the chapter.

As for Marcus, his biggest fear is that AI could "get trapped in a local minimum, dwelling too heavily in the wrong part of intellectual space, focusing too much on the detailed exploration of a particular class of accessible but limited models that are geared around capturing low-hanging fruit—potentially neglecting riskier excursions that might ultimately lead to a more robust path."

However, he is not a pessimist. He believes that researchers need to go beyond deep learning and find new techniques that can solve tough problems.

Conclusion

While Marcus has pointed out the flaws in deep learning, the fact is that this AI approach is still extremely powerful. In less than a decade, it has revolutionized the tech world—and is also significantly impacting areas like finance, robotics, and healthcare.

With the surge in investments from large tech companies and VCs, there will be further innovation with the models. This will also encourage engineers to get postgraduate degrees, creating a virtuous cycle of breakthroughs.

Key Takeaways

- Deep learning, which is a subfield of machine learning, processes huge amounts of data to detect relationships and patterns that humans are often unable to detect. The word "deep" describes the number of hidden layers.
- An artificial neural network (ANN) is a function that includes units that have weights and are used to predict values in an AI model.
- A hidden layer is a part of a model that processes incoming data.
- A feed-forward neural network has data that goes only from input to the hidden layer to the output. The results do not cycle back. Yet they can go into another neural network.
- An activation function is non-linear. In other words, it tends to do a better job of reflecting the real world.
- A sigmoid is an activation function that compresses the input value into a range of 0–1, which makes it easier for analysis.
- Backpropagation is a sophisticated technique to adjust the weights in a neural network. This approach has been critical for the growth in deep learning.
- A recurrent neural network (RNN) is a function that not only processes the input but also prior inputs across time.

- A convolutional neural network (CNN) analyzes data section by section (that is, by convolutions). This model is geared for complex applications like image recognition.
- A generative adversarial network or GAN is where two neural networks compete with each other in a tight feedback loop. The result is often the creation of a new object.
- Explainability describes techniques for transparency with complex deep learning models.

CHAPTER

5

Robotic Process Automation (RPA)

An Easier Path to AI

By interacting with applications just as a human would, software robots can open email attachments, complete e-forms, record and re-key data, and perform other tasks that mimic human action.

—Kaushik Iyengar,
director of Digital Transformation and Optimization at AT&T[1]

Back in 2005, Daniel Dines and Marius Tirca founded UiPath, which was located in Bucharest, Romania. The company focused mostly on providing integration services for applications from Google, Microsoft, and IBM. But it was a struggle as the company relied mostly on custom work for clients.

[1] www2.deloitte.com/insights/us/en/focus/signals-for-strategists/cognitive-enterprise-robotic-process-automation.html