CHAPTER 12

Democratization and Impacts of AI

> *A diversity of thought, perspective and culture is important in any field.*
>
> —Sarah Friar, chief financial officer at Square

Machine learning is shifting from being a tailor-made service of custom-crafted models into an era of productization. The new class of products and services coming out of this shift allows a diverse set of people to train their own models with their own data without the help of machine learning researchers. The tools are changing from being available only to large corporations that can afford it into the hands of smaller businesses that may be struggling to compete. As the reliance on specialists decreases for many use cases, it will become commonplace to see businesses of all sizes relying on artificial intelligence to improve some aspect of their daily operations. While the long-term impacts of this change have yet to be proven, the rise of AI automation tools has stirred all kinds of debate, especially around economics and jobs.

In terms of democratization, we still have a long way to go. However, we are making strides in increasing the accessibility of AI. This year may mark a milestone in the machine learning ecosystem. In January 2018, Gartner predicted that by 2019, organizations using self-service analytics and business intelligence tools would output more analytics than

L. Luce, *Artificial Intelligence for Fashion*,
https://doi.org/10.1007/978-1-4842-3931-5_12

professional analysts. While this statistic encompasses a wider breadth of technologies, AI plays a major role. The topic of the democratization of AI could not be more of the moment.

Software accessibility isn't the only driving force behind this shift. Access to data, the commoditization of specialized hardware, and an up-cropping of platforms to run that hardware are also playing their part.

Lowering the Barrier to Entry

It might not be obvious to professionals working outside the tech industry, but until recently, even experienced software developers might have had a difficult time implementing machine learning techniques. They might not have the background in mathematics or the domain knowledge to get machine learning models up and running. They also simply might not have had a reason to do it.

The progress that has been made in this field in the last five years has changed this dramatically. Every year, new tools are being introduced by the **open source** community, some of them with wide adoption. Products that are entering the market, especially **cloud services**, have also made it easier to train and run models.

Simplified Interfaces

As described in Chapter 7, there is a rise of increasingly powerful graphical interfaces for hosting data, training, and running models. With these GUIs, more can be done than ever before without writing code.

In terms of ease of use for the end user, solutions built for enterprise are built to have a more user-friendly GUI and automation in the back end, with the goal to increase the speed of execution. These platforms are built to make it easy for key stakeholders to understand and for scientists to implement.

One example of this is **DataRobot**, which offers enterprise products for running machine learning models. It's likely if you're using DataRobot that you have a data science team that can run and interpret experiments on the platform.

The goal of these platforms is to be able to do more faster, with little to no coding. It's not that these platforms make it easy for anyone at all to start using, but that for those who know the core concepts and how to use them, enterprise tools like DataRobot can provide powerful self-service automation.

Developer Tools

There are also tools accessible for developers looking to start implementing machine learning in the products they're working on.

Google's **Machine Learning Kit** (**ML Kit**) makes commonly used machine learning techniques available and production ready for developer use. They provide easy access to get started with these tools through their mobile and web app hosting service, Firebase.

ML Kit works with Google's **TensorFlow** Lite. TensorFlow is an open source software library for machine learning problems. It's meant to be able to run on everyday computers rather than just specialized machines. It also gives researchers a jump start at implementing neural networks. TensorFlow Lite is a version of this that is optimized for even lighter computation commonly required to run on mobile devices.

Access to Data

In March 2017, Google announced its acquisition of Kaggle, a data science platform mentioned in Chapter 7. Kaggle contains one of the largest repositories of datasets; an invaluable tool for learning and practicing machine learning. Before the acquisition, Google started releasing large labeled datasets to the platform, like a labeled set of YouTube videos. Access to this kind of resource was something that only large companies had before. Giving more people access to high-quality real-world data accelerates the rate of growth in the industry. For Google, this is a win, because when people can have better resources to learn from, they gain access to a larger pool of talent and help grow the market for their own services.

Kaggle isn't the only resource that people can go to for large datasets. Corporations like Amazon also provide lists of public datasets that are maintained by third parties and hosted on their platforms. In Amazon's case, this is through **Amazon Web Services** (**AWS**). Universities and government agencies frequently provide public data. New York–based startup Vigilant provides easy access to thousands of public records like these.

In September 2018, Google released Google Dataset Search, a tool to make it easy to search the Web for datasets: `https://toolbox.google.com/datasetsearch`.

Open Source

The word *open source* has been brought up a few times already in this chapter. Open source refers to software that has been made publicly available and can be modified or shared. Machine learning has progressed rapidly in part because of the open sourcing of tools and research.

Researchers have long been publishing academic papers containing their machine learning findings and the algorithms they used to achieve those results. What's different in the last five to ten years is that those papers are now linked to online media including videos, photos, repositories that host the code to run the algorithms, and even sometimes to demos that show you how they work or let you run them with your own data. Before, demos that ran neural networks and other complex models would not have been possible to host on the Web.

Specialized Hardware

Getting to where we are today with artificial intelligence has required the emergence of an entire ecosystem for machine learning to develop. Without computing power, AI processes like deep learning would not be possible. Until recently, the amount of computation required to run these models was more than a standard computer could handle.

GPUs and TPUs

Before the introduction of **graphics processing units** (**GPUs**) for machine learning in the early 2000s, neural networks were not feasible for production or even for research purposes. A GPU is similar to a **central processing unit** (**CPU**), the electronic circuit that carries out instructions for your computer to work, but it can run many more computations at the same time than a CPU.

GPUs were originally designed for video games. In order to render graphics on a screen, video games run many complex calculations at once to quickly determine the value of every pixel on the screen based on what's happening in the game. Training neural networks and other machine learning models similarly requires a lot of computation. GPUs allow computers to run those computations in parallel, reducing the overall time needed to train. According to OpenAI, the amount of compute power used for training large jobs has doubled every 3.5 months since 2012.

Before 2012, it was very uncommon to use GPUs for machine learning. Between 2012 and 2014, infrastructure was built to make training on GPUs possible.

Around 2016, some of the approaches changed so that processes could run in parallel, and **tensor processing units** (**TPUs**) were introduced to the mix. The hardware has been constantly evolving to keep up with the demands for faster and more complex computation.

TPUs were introduced by Google. They are specialized hardware for deep learning, optimized for the basic operations required to train a neural network. The release of the TPU announcement garnered excitement in the industry because of a 15 to 30 times improvement in performance using these chips and even higher returns in terms of performance per watt. While these chips aren't available for purchase themselves, you can essentially rent them through Google's cloud computing services. Figure 12-1 shows a TPU circuit board and Google server center.

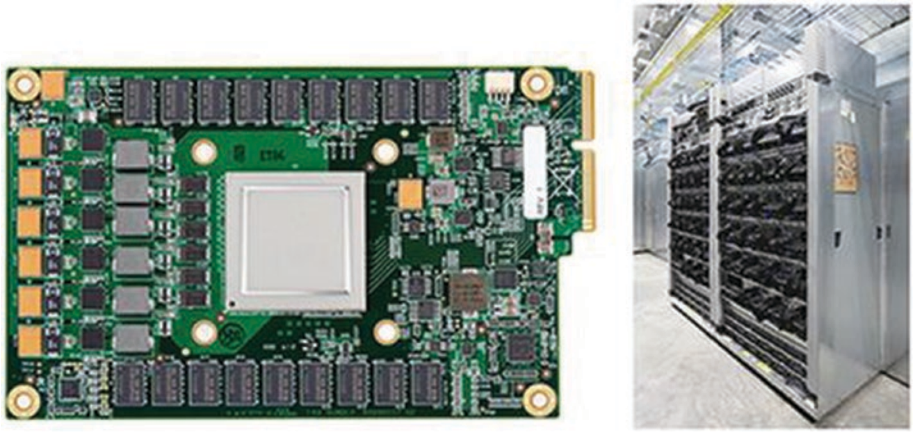

Figure 12-1. *Tensor processing units (left) and their server center (right), image courtesy of Google.*

GPUs and TPUs aren't the full story. There have been a number of startups also building in the space. According to the *New York Times*, as of 2018, at least 45 startups were working on building machine-learning specialized chips alone.

Cloud Services

Even more recently, cloud computing using GPUs has become a viable solution for businesses looking to implement machine learning models without purchasing their own specialized hardware. While today, this is still expensive to maintain, new algorithms in development help optimize the use of expensive server time, reducing the costs significantly.

Cloud GPUs give experts the ability to run complex models on rented servers. FloydHub, Google Cloud GPUs, and Amazon ML all provide access to cloud computing of this type.

In January 2018, Google announced a new web service, **Cloud AutoML**, which allows businesses to use a number of their models without an in-house expert. Customers can complete tasks that are relevant to their businesses, like image classification, natural language processing, and language translation. Google plans to expand into other areas and offer a suite of easy-to-use machine learning tools.

Tutorials and Online Courses

In online education marketplaces like **Udemy** or **Coursera**, machine learning and related courses are increasing in prevalence. As someone with little technical experience, you can take one of these courses and get a step-by-step walk-through on machine learning. For example, you can learn how to build your own recommender system or train a neural network using popular programming languages.

Besides these marketplaces, tutorials in the space are easy to come across all over the Internet, on the blogs of startups and researchers.

Impact on Jobs

> *It will not be a world of man versus machine, it will be a world of man plus machine.*
>
> —Virginia Rometty, chair, president, and CEO, IBM

Artificial intelligence is often feared as a technology that will replace workers by automating their jobs. AI is good at automation and even outperforms humans at certain tasks, but not every task required to do most jobs. Throughout history, there have been more examples of automation creating jobs than taking them away.

The industrial revolution increased the amount of cloth a single weaver could produce by 50 times and decreased the labor per yard by 98%. This sounds like it would cause a near-elimination of workers in the industry, but in practice the inverse happened. With the reduction in the price of cloth came a steep increase in demand, creating four times more jobs.

Will this time be different? Technological change today is happening at a much quicker pace than it did in the 19th century. While the question remains impossible to answer with absolute confidence, many experts have weighed in with doubts that AI will be replacing entire jobs and industries anytime soon. However, some industries and the countries reliant on those industries will be affected more than others.

Ethics and the Future

> *Is artificial intelligence less than our intelligence?*
>
> —Spike Jonze, filmmaker

As the field of AI has matured, issues around its ethics have been called into question. Not only has it become more difficult to understand why a model makes a particular choice, but there has been little accountability in terms of understanding the biases inherent in the data they're trained on. This ambiguity has spurred a lot of criticism. How can we ensure that we aren't propagating our human biases when applying this new technology?

Race and Gender

A few alarming examples of AI biases about race and gender have sparked ethical conversations around discrimination. In early 2018, a study conducted by researcher Joy Buolamwini at the MIT Media Lab revealed starkly different error rates based on skin color and gender in facial recognition systems. In one of her examples, she shows that the accuracy of these systems was disproportionately more accurate for white men, with a 1% error rate in identifying gender, than it was for black women, who had up to 35% error rate.

At least in part, these biases lie in the data being passed through ML models to train them. As a potential step toward resolving these biases, Timnit Gebru and her coauthors from Microsoft Research proposed Datasheets for Datasets. The idea is that every dataset comes with a sort of nutrition label that explains details such as who made the dataset, where the data came from, and how it was created.

In all of computer science, it is estimated that women hold an estimated 13% of jobs. I personally believe that it will be difficult to resolve issues of bias without addressing issues of representation.

The Partnership on AI

The controversies around ethics expand beyond discrimination. Autonomous drone warfare and other war-related AI technologies also lie in a gray area. At what point do we begin introducing regulations around AI? Will regulation stunt progress in this field?

In January 2017, the Partnership on AI was formed by major technology companies to create open dialogue about ethics and implications of AI technology development. Since then, over 70 organizations and corporations worldwide have joined the nonprofit partnership. Their core tenants are around building technology that we understand.

The future really just depends on how we choose to use these powerful new tools.

Summary

The surge in new tools and hardware being introduced to the market has made artificial intelligence more accessible now than ever been before. Increases in compute power, open source machine learning models, specialized hardware, and cloud services specialized for the AI industry have all contributed to that growth.

The popularity and power of this technology has called a lot into question. Will it crash the job market? Is it ethical? We cannot predict what will happen in the future, but the technology itself will only do what the people controlling it ask it to do. We are a long way from AI taking on a mind of its own and destroying humanity; that discourse is a distraction from holding the humans who built it accountable.

Terminology from This Chapter

Amazon Web Services (**AWS**)—A suite of cloud computing services offered by Amazon on a paid subscription basis. See also *cloud services.*

Central processing units (**CPUs**)—A standard processor that exists in most consumer computers today. These are the "brains" of the computer, where operations are processed.

Cloud AutoML—A product offered by Google that makes it possible to train machine learning models with little machine learning expertise.

Cloud GPUs—Cloud computing services that specifically offer use of GPUs.

Cloud services—Broadly include services that are hosted over the Internet. Using cloud services, businesses or individuals don't need to maintain their own servers in order to offer or use certain services. In the context of machine learning, not needing to own a computer with a lot of processing power is very convenient.

Coursera—An online learning platform that offers not only courses in specialized topics, but also degrees. The platform works with top institutions like Yale and Stanford, as well as top companies like Google and IBM to offer high-quality courses.

DataRobot—A machine learning platform that automates some of the processes required to build and deploy machine learning models. It offers an easy-to-use GUI that makes it possible to implement these models without writing code.

Graphics processing units (**GPUs**)—A specialized electronic component often used in computers that require a lot of image processing or graphics rendering, as in video games. In the early 2000s, GPUs found their way into machine learning because of their higher processing power, which could be utilized in challenging computational problems.

Machine Learning Kit (**ML Kit**)—A product created by Google that makes it easy to implement certain machine learning products during application development.

Open source—Software in which the source code and the algorithms it is implementing are publicly accessible.

TensorFlow—Created by the Google Brain team at Google and released as open source in 2015. It is a software library primarily used for machine learning. TensorFlow Lite is an adaptation of TensorFlow that is built to run on Android devices, which have significantly less compute power than larger machines.

Tensor processing units (**TPUs**)—A hardware component introduced by Google to speed up computation during model training.

Udemy—An online education platform in which content generators can earn money in return for their courses. It is not a traditional academic program, and you cannot get college credit for taking courses there.