1

Machine Learning—The Early Years

Confucius reportedly said: *"Learn avidly. Question repeatedly what you have learned. Analyze it carefully. Then put what you have learned into practice intelligently."* We thought this quote was very apt as we prepared to write this book on artificial intelligence (AI) and machine learning (ML).

Revolution: Change Is Inevitable

Throughout the ages, when man has introduced machines to help automate processes, improve productivity, or reduce overheads, segments of society have rejected their introduction.

During the industrial revolution, spanning from the 1700s through the 1800s, the cotton gin and steam power, for example, helped create and automate textile mills across the United Kingdom, Europe, and the United States of America. This period saw a big change in culture and society. Workers moved from the country and farms to work in cities. Factories were positioned close to rivers and waterways to power water wheels and leverage transport routes. Transportation infrastructure grew and blossomed as part of the supply chain. Printers were able to harness steam power to print newspapers and books cheaply, which helped more people get the news and learn how to read.

Yet, some people felt threatened by this change—fearful for their futures. Humankind often has mixed feelings about change—even if there are clear benefits—sometimes to the point of resistance and sabotage. For instance, a group of weavers in Britain who lost their jobs to large factories began to fight back by rioting and destroying machinery. (These individuals, led by Ned Ludd, became known as "Luddites.")

The Computing Revolution

Despite the many advances and benefits—from simple automation to space exploration, from better understanding of genetics and DNA sequencing to new treatments for chronic illness—the computing industry has also faced resistance. Some of this opposition has been based on fear, uncertainty, and doubt (FUD). Often, that's because we find ourselves in uncharted territories, where our established legal systems, ethical systems, doctrines—even our personal value systems—may be challenged. New technologies and discoveries can suddenly demonstrate that what we once thought impossible has become achievable, as well as what we deemed to be true to be untrue (and vice versa).

From Number Crunching to Analytics

Since the 1960s, computers have increasingly demonstrated their ability to help automate and repeat complex tasks faster and with more accuracy than humans alone. For example, they are frequently used to perform complex calculations, bank transactions, stock trading, airline reservations, space exploration, and so forth.

We are often told that we should "learn from our mistakes." Yet history demonstrates that, despite our best efforts, we repeat those mistakes, sometimes to the detriment of our planet and our very existence. We would not want banking applications, health monitoring, or defense and safety systems to fail to detect an error that would leave us vulnerable over and over again. It would be advantageous if those systems could continually learn and become smarter with each interaction. Indeed, machine learning (ML), a branch of the larger artificial intelligence (AI) discipline and a part of the analytics landscape, can help do just that.

Jargon and Buzzwords

Big data. *Algorithms*. *Models*. *Machine learning*. *Data science*. *Neural networks*. *Natural language processing*. *Knowledge representation*. *Decision optimization*. We read and hear these buzzwords often these days. They are all part of what is known in the industry as AI and cognitive computing.

You can think of machine learning (ML) as a collection of software components (libraries) and an execution engine for running a set of algorithms as part of a model to predict one or more outcomes. Each outcome will have an associated score that indicates the confidence level at which it will occur. If you have studied statistics, this should sound familiar. Cognitive computing is the ability of computers to simulate the human behavior of understanding, reasoning, and thought processing. It's data-hungry and leverages many forms of machine learning (ML), including vision, natural language processing, speech, sentiment analysis, and other forms of analytics. The ultimate goal is to simulate intelligence though a set of software and hardware services to produce better business outcomes. Hence the term "artificial intelligence" (AI), as shown in Figure 1.1.

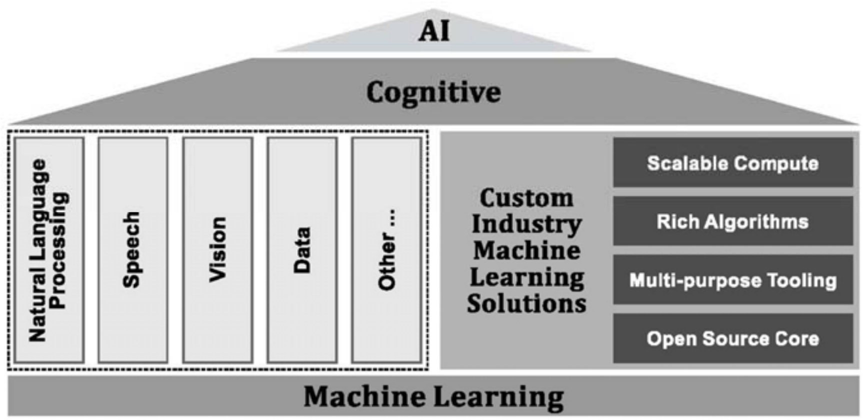

Figure 1.1: Positioning ML, cognitive, and AI

From Checkers to *Jeopardy!*

The phrase "machine learning" is said to have been coined by an IBMer named Arthur Samuel in 1959, even though his earlier research on the subject started back in 1949. It is defined as *a field of computer science that gives computers the ability to learn without being explicitly programmed*. Samuel is known within the AI community for his groundbreaking work in teaching a computer to play checkers (Figure 1.2). He implemented what is now known as "alpha-beta pruning." Instead of searching each path until it came to the game's conclusion, Samuel developed a scoring function based on the position of the board at any given time. This function tried to measure the chance of winning for each side at the given position, taking into account many game factors. Using a “minimax” strategy, the program made the move that optimized the value of this function.

In what he called “rote learning,” the program remembered every position it had already seen, along with the terminal value (often the maximized value) of the reward function.

Figure 1.2: Arthur Samuel of IBM demonstrating computer checkers game, 1959

(Source: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/ibm700series/impacts/*)*

Machine learning grew out of the quest for AI. Even in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed “neural networks.” These were mostly “*perceptrons*” (which will be explained shortly) and other models that were later found to be reinventions of the generalized linear statistical model. Probabilistic reasoning was also employed, especially in automated medical diagnosis.

Two important learning concepts need to be introduced here and will be expanded on later:

- **Supervised Learning:** An algorithm is given training data that contains the “correct answer” for each example. For instance, a supervised learning algorithm for credit-card fraud detection would take as input a set of recorded transactions. For each transaction, the training data would contain a flag that says whether it is fraudulent or not.

- **Unsupervised Learning:** An algorithm looks for structure in the training data, such as finding which examples are similar to each other, and then groups them in clusters.

Around 1957, the perceptron was conceived. It is an algorithm for supervised learning or binary classifiers (that is, functions that can decide whether an input, represented by a vector of numbers, belongs to some specific class or not). Its initial implementation (in customized hardware) was one of the first artificial neural networks to be produced.

However, an increasing emphasis on the logical, knowledge-based approach caused a rift between AI and machine-learning experts. Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation. By 1980, “expert systems” (computer systems that emulate the decision-making ability of a human expert) had come to dominate AI, and statistics was out of favor. Work on symbolic/knowledge-based learning continued within AI, but the more statistical line of research was now outside the field of AI and was being used in pattern recognition and information retrieval. Neural network research had been abandoned by AI and computer science around the same time.

During the 1970s and 1980s, in what is now considered the “AI winter,” pessimism abounded in the AI community, which was reflected by the press. The result was severe cutbacks in funding, followed by the end of serious research. The message to the business community was that enthusiasm for AI had spiraled out of control in the ’80s and disappointment would certainly follow. Three years later, the billion-dollar AI industry began to collapse.

Later in the 1980s, a rediscovery of “backpropagation” caused a resurgence in machine learning research. This was followed in the 1990s by a shift of ML from a knowledge-driven approach to a data-driven approach. Scientists began creating programs for computers to analyze large amounts of data and draw conclusions or “learn” from the results. Recurrent neural networks also became popular. From 2000 to 2010, kernel methods (a class of algorithms used for pattern analysis, whose best-known member is the support vector machine) grew in popularity, and competitive machine learning became more widespread.

By 2010, deep learning (DL) was helping machine learning become integral to many widely used software services and applications. Deep learning uses more advanced techniques, such as convolutional neural networks, which look at things more deeply—in layers. Deep learning architectures such as deep neural networks, deep belief networks, and recurrent neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social-network filtering, machine translation, bioinformatics, and drug design, where they have produced results comparable to and in some cases superior to human experts.

In 2011, IBM's "Watson" AI system competed on the *Jeopardy!* TV game, beating two top contestants. It had access to 200 million pages of structured and unstructured content, consuming terabytes of disk storage, including the full text of Wikipedia (but was not connected to the Internet during the game). For each clue, Watson's three most probable responses were displayed on the television screen. Watson consistently outperformed its human opponents on the game's signaling device, but had trouble in a few categories, notably those having short clues containing only a few words—kind of similar to human language ambiguity. Its natural language processing, predictive scoring, and models were key to its success.

The Impact of Big Data

There are a lot of definitions for "big data," but let us set the record straight. Big data just means *all* data: structured data in databases; log files; data in documents; data held in video, voice, and image files; social-media feeds; and more. The data can be static or in motion. Just remember the 3 *V*s : volume, variety, velocity.

Big data is important because the more data (volume and variety) we have, the more informed our insights should be. Trying to make a decision on limited data can incur risk because we may not have the full picture. For example, making a decision on how to promote a product based solely on a customer's transactions might be risky. By including customer sentiment, previous product reviews, and aspects of an individual's social life, hobbies, and lifestyle, we should be able to more accurately understand the needs and wants of the individual.

When we apply AI techniques to big data, the result should be smarter business outcomes because we are considering interrelationships across many different data sources. Herein lies a big challenge for many organizations that have over the years accumulated data that was never designed to be shared. Another challenge is whether or not organizations have the correct information architecture to support their artificial intelligence initiatives.

Cognitive Computing Comes of Age

Around 2015, there was a convergence of many facets of ML and deep learning (including but not limited to natural language processing, language interpretation, visual and voice recognition, translation, and sentiment analysis technologies) into what is known as cognitive computing. As stated earlier, cognitive computing is the ability of computers to simulate human behavior of understanding, reasoning, and thought processing. This ability, combined with open source, improved tools, the demand for self-service "pay as you go" analytics, cheap compute power, the ability to ingest massive amounts of any kind of data, scale-out processing across on-premise, and private and public clouds helped democratize cognitive computing. Finally, it was within the grasp of the vast majority of data scientists, data engineers, developers, and ad hoc analysts to deliver and consume AI capabilities. Watson had showcased many of these capabilities a few years earlier on the *Jeopardy!* TV game.

Since the *Jeopardy!* game, AI has been applied across many industries—from financial markets in order to help predict and prevent fraud in real time, to retail in an effort to help predict what customers will purchase next, to security and protection in an attempt to help prevent cyber-attacks and crimes, to media with an aim to help tailor the viewing experience with bespoke advertising, to meteorology for predicting weather and climate change, and to healthcare in order to help clinicians design cancer treatment plans faster and with more success.

Who Is a Data Scientist?

In November 1997, C.F. Jeff Wu gave a lecture entitled "Statistics = Data Science?," where he characterized statistical work as a trilogy of data collection, data modeling and analysis, and decision-making. He advocated that statistics be renamed "data science" and statisticians "data scientists." In 2012, *Harvard Business Review* called it "The Sexiest Job of the 21st Century." This hybrid role demanded a lot of specific skills in the areas described above, which in turn made high-quality data scientists difficult to find and retain.

While also involved in some aspects of data science, data engineers are more focused on the infrastructure and information architecture necessary to support data science projects, as shown in Figure 1.3.

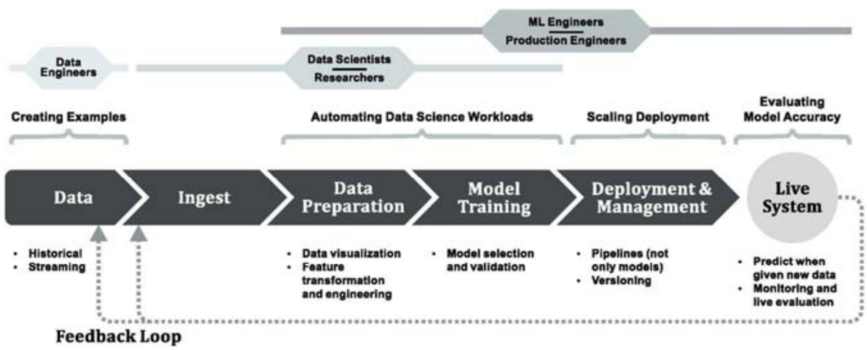

Figure 1.3: A machine-learning flow with persona involvement

ML, AI, Deep Learning, Cognitive Computing, Data Science—Positioning Summary

These terms are often used interchangeably and mean different things to different people. While they may be open to interpretation, this book provides an example of their relationships with each other.

- Artificial intelligence (AI) is the broad concept of machines being able to carry out tasks in a way that we would consider “smart.” The IBM Watson win at *Jeopardy!* and the Google AlphaGo win against top Go players are good examples of significant AI achievements. Reasoning, planning, and natural language understanding are AI capabilities.
- Machine learning (ML) is an AI capability based around the idea that we should be able to give machines access to data and let them learn for themselves.
- Decision optimization (DO) is a form of machine learning that analyzes multiple parameters and variables to help people and applications take prescriptive actions based on what is deemed to be the optimized decision.

- Deep learning (DL) is a class of machine-learning algorithms that produces models inspired by the human brain: neural networks. These algorithms are used for processing unstructured data such as images, sound, and natural language.
- Cognitive systems are systems that have machine-learning and other AI capabilities such as natural language processing and audio and speech recognition embedded in them to help simulate human behavior of reasoning, understanding, and thought processing.
- Data science is the practice of building models from data—a key piece of the machine-learning process.