

Introduction

The world is changing. There are more phones than people in the world, and it is increasingly connected. People use virtual assistants, self-driving cars, find partners through digital apps, and search the Web for any symptom of ill health. Each digital event leaves a digital exhaust that is datafying life as we know it. The success of many of the world's most loved services, from Google to Uber, Alexa to Netflix, is grounded in big data and optimization.

Although medicine has been receptive to the benefits of big data and AI, it has been slow to adopt the rapidly evolving technology, particularly when compared to sectors such as finance, entertainment, and transport: that is, until now.

Recent digital disruption has catalyzed healthcare's adoption of big data and AI. Data of all sorts, shapes, and sizes are used to train AI technologies that facilitate machines to learn, adapt, and improve on their learning. Academic institutions and start-ups alike are developing rapid prototype technologies with increasingly robust health and engagement claims. The blending of technology and medicine has been expedited by smartphones and the Internet of Things, which is facilitating a wealth of innovation that continues to improve lives. With the arrival of health technology, people can monitor their health without the assistance of a healthcare professional; healthcare is now mobile and no longer in the waiting room.

At the same time, the world's population is living longer, and unhealthier, than ever: and in a financial crisis. Healthcare services are turning to value-based and incentivized care as non-communicable diseases such as type 2 diabetes and obesity become global pandemics.

INTRODUCTION

Data is proving an invaluable tool in improving health; information is empowering. Data science, the science of big data, its analysis, and intelligent programming layers has now become a pillar to achieve traction and success in healthcare. Digital health can democratize and personalize healthcare—and data is the golden key. This comes at the same time as a growing appetite to measure and quantify more aspects of human life.

Data insight and real-world evidence are facilitating rapid technology innovation that regulators are struggling to keep up with. The consequences of digital health democratization have not only a health impact but also ethical implications. Big data and machine learning enable stakeholders to uncover hidden connections and patterns, including predictions of the future. The effects of understanding this data have moral and legal consequences that require appropriate governance to mitigate risk and harm.

Healthcare providers, individuals, and organizations all house a fountain of data that can be used for machine learning. Many people have a fair idea of what they would like to learn from data but are unaware as to how much data is required and what can be achieved before more technical aspects of uncovering hidden patterns, trends, and biases are found.

This book takes a practical, hands-on approach to big data, AI, and machine learning and the ethical implications of such tools. We cover the theory and practical applications of AI in healthcare—from where and how to start to applying machine learning techniques and evaluating performance. The book concludes with a series of case studies from leading health organizations who utilize AI and big data in novel and innovative ways.

CHAPTER 1

What Is Artificial Intelligence?

“Knowledge on its own is nothing, but the application of useful knowledge? That’s powerful.”

—Osho

Artificial intelligence (AI) is considered, once again, to be one of the most exciting advances of our time. Virtual assistants can determine our music tastes with remarkable accuracy, cars are now able to drive themselves, and mobile apps can reverse diseases once considered to be chronic and progressive.

Many people are surprised to learn that AI is nothing new. AI technologies have existed for decades. It is, in fact, going through a resurgence—and it is being driven by availability of data and cheaper computing.

A Multifaceted Discipline

AI is a subset of computer science that has origins in mathematics, logic, philosophy, psychology, cognitive science, and biology, among others (Figure 1-1).

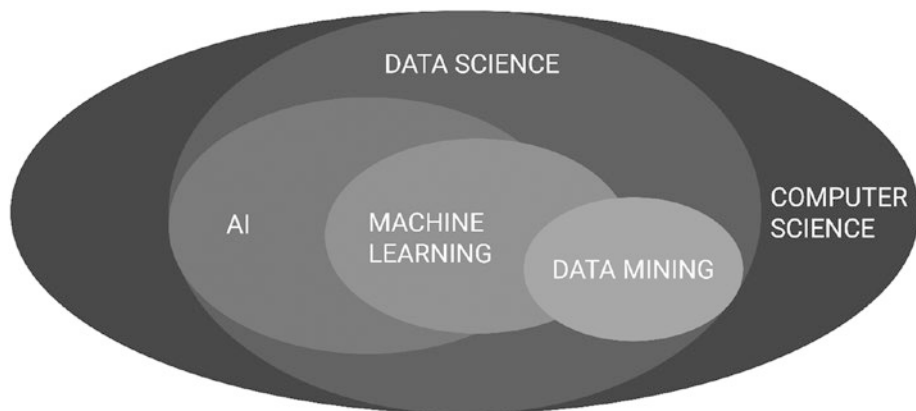


Figure 1-1. *AI, machine learning, and their place in computer science*

The earliest research into AI was inspired by a constellation of thought that began in the late 1930s and culminated in 1950 when British pioneer Alan Turing published *Computing Machinery and Intelligence* in which he asked, can machines think? The Turing Test proposed a test of a machine's ability to demonstrate “artificial” intelligence, evaluating whether the behavior of a machine is indistinguishable from that of a human. Turing proposed that a computer could be considered to be able to think if a human evaluator could have a natural language conversation with both a computer and a human and not distinguish between either (i.e., an agent or system that is successfully mimicking human behavior).

The term AI was first coined in 1956 by Professor John McCarthy of Dartmouth College. Professor McCarthy proposed a summer research project based on the idea that “every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it”.[1]

The truth is that AI, at its core, is merely programming. As depicted in Figure 1-1, AI can be understood as an abstraction of computer science. The surge in its popularity, and so too its ability, has much to do with the

explosion of data through mobile devices, smartwatches, wearables, and the ability to access computer power cheaper than ever before. It was estimated by IBM in 2011 that 90% of global data had been created in the preceding 2 years.[2]

It is estimated that there will be 150 billion networked measuring sensors in the next decade—which is 20 times the global population. This exponential data generated is enabling everything to become smart. From smartphones to smart washing cars, smart homes, cities, and communities await.

With this data comes a plethora of learning opportunities and hence, the focus has now shifted to learning from available data and the development of intelligent systems. The more data a system is given, the more it is capable of learning, which allows it to become more accurate.

The use and application of AI and machine learning in enterprise are still relatively new, and even more so in health. The Gartner “Hype Cycle for Emerging Technologies” in 2017 placed machine learning in the peak of inflated expectations, with 5 to 10 years before plateau.

As a result, the applications of machine learning within the healthcare setting are fresh, exciting, and innovative. With more mobile devices than people today, the future of health is wrought with data from the patient, environment, and physician. As a result, the opportunity for optimizing health with AI and machine learning is ripening.

The realization of AI and machine learning in health could not be more welcome in the current ecosystem, as healthcare costs are increasing globally, and governmental and private bill payers are placing increasing pressures on services to become more cost-effective. Costs must typically be managed without negatively impacting patient access, patient care, and health outcomes.

But how can AI and machine learning be applied in an everyday healthcare setting? This book is intended for those who seek to understand what AI and machine learning are and how intelligent systems can be

developed, evaluated, and deployed within their health ecosystem. Real-life case studies in health intelligence are included, with examples of how AI and machine learning is improving patient health, population health, and facilitating significant cost savings and efficiencies.

By the end of the book, readers should be confident in explaining key aspects of AI and machine learning to stakeholders. Readers will be able to describe the machine learning approach and limitations, fundamental algorithms, the usefulness and requirements of data, the ethics and governance of learning, and how to evaluate the success of such systems.

Rather than focus on overwhelming statistics and algebra, theory and practical applications of AI and machine learning in healthcare are explored—with methods and tips on how to evaluate the efficacy, suitability, and success of AI and machine learning applications.

Examining Artificial Intelligence

At its heart, AI can be defined as the simulation of intelligent behavior in agents (computers) in a manner that we, as humans, would consider to be smart or human-like. The core concepts of AI include agents developing traits including knowledge, reasoning, problem-solving, perception, learning, planning, and the ability to manipulate and move.

In particular, AI could be considered to comprise the following:

- Getting a system to reason rationally. Techniques include automated reasoning, proof planning, constraint solving, and case-based reasoning.
- Getting a program to learn, discover and predict. Techniques include machine learning, data mining (search), and scientific knowledge discovery.

- Getting a program to play games. Techniques include minimax search and alpha-beta pruning.
- Getting a program to communicate with humans. Techniques include natural language processing (NLP).
- Getting a program to exhibit signs of life. Techniques include genetic algorithms.
- Enabling machines to navigate intelligently in the world
- This involves robotic techniques such as planning and vision.

There are many misconceptions of AI, primarily as it's still quite a young discipline. Indeed, there are also many views as to how it will develop. Interesting expert opinions include those of Kevin Warwick, who is of the opinion robots will take over the earth. Roger Penrose reasons that computers can never truly be intelligent. Meanwhile, Mark Jeffery goes as far as to suggest that computers will evolve to be human. Whether AI will take over the earth in the next generation is unlikely, but AI and its applications are here to stay.

In the past, the intelligence aspect of AI has been stunted due to limited datasets, representative samples of data, and the inability to both store and subsequently index and analyze considerable volumes of data. Today, data comes in real time, fuelled by exponential growth in mobile phone usage, digital devices, increasingly digitized systems, wearables, and the Internet of Things (IoT).

Not only is data now streaming in real time, but it also comes in at a rapid pace, from a variety of sources, and with the demand that it must be available for analysis, and fundamentally interpretable, to make better decisions.

There are four distinctive categories of AI.

Reactive Machines

This is the most basic AI. Reactive systems respond in a current scenario, relying on taught or recalled data to make decisions in their current state. Reactive machines perform the tasks they are designed for well, but they can do nothing else. This is because these systems are not able to use past experiences to affect future decisions. This does not mean reactive machines are useless. Deep Blue, the chess-playing IBM supercomputer, was a reactive machine, able to make predictions based on the chess board at that point in time. Deep Blue beat world champion chess player Garry Kasparov in 1996. A little-known fact is that Kasparov won three of the remaining five games and defeated Deep Blue by four games to two.

Limited Memory—Systems That Think and Act Rationally

AI that works off the principle of limited memory and uses both pre-programmed knowledge and subsequent observations carried out over time. During observations, the system looks at items within its environment and detects how they change, then makes necessary adjustments. This technology is used in autonomous cars. Ubiquitous Internet access and IoT is providing an infinite source of knowledge for limited memory systems.

Theory of Mind—Systems That Think Like Humans

Theory of mind AI represents systems that interpret their worlds and the actors, or people, in them. This kind of AI requires an understanding that the people and things within an environment can also alter their feelings and behaviors. Although such AI is presently limited, it could be used in caregiving roles such as assisting elderly or disabled people with everyday tasks.

As such, a robot that is working with a theory of mind AI would be able to gauge things within their worlds and recognize that the people within the environments have their own minds, unique emotions, learned experiences, and so on. Theory of mind AI can attempt to understand people's intentions and predict how they may behave.

Self-Aware AI—Systems That Are Humans

This most advanced type of AI involves machines that have consciousness and recognize the world beyond humans. This AI does not exist yet, but software has been demonstrated with desires for certain things and recognition of its own internal feelings. For instance, in 2015, researchers at the Rensselaer Polytechnic Institute gave an updated version of the wise men puzzle, an induction self-awareness test, to three robots—and one passed. The test requires AI to listen and understand unstructured text as well as being able to recognize its own voice and its distinction from other robots.

Technology is now agile enough to access huge datasets in real time, and learn on the go. Ultimately, AI is only as good as the data that's used to create it—and with robust, high-volume data, we can be more confident about our decisions.

Healthcare has been slow to adopt the benefits of big data and AI, especially when compared to transport, energy, and finance. Although there are many reasons for this, the rigidity of the medical health sector has been duly grounded in the fact that peoples' lives are at risk. Medical services are more of a necessity than a consumer choice; so historically, the medical industry has had little to no threat that usually drives other industries to seek innovation.

That has expedited a gap in what healthcare institutes can provide and what patients want—which subsequently has led to variances in care, in health outcomes, and a globally recognized medication-first approach to disease.

The explosion of data has propelled AI through enabling a data-led approach to intelligence. The last 5 years has been particularly disruptive in healthcare, with applications of data-led AI helping intelligent systems not only to predict, diagnose, and manage disease but to actively reverse and prevent it—and the realization of digital therapeutics.

Recent advances in image recognition and classification are beginning to filter into industry too, with deep neural networks achieving remarkable success in visual recognition tasks, often matching or exceeding human performance (Figure 1-2).

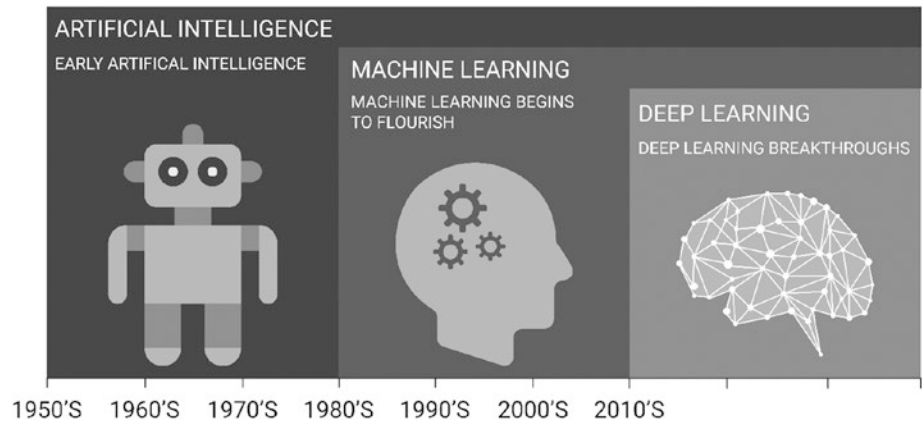


Figure 1-2. *AI and its development*

What Is Machine Learning?

Machine learning is a term credited to Arthur Samuel of IBM, who in 1959 proposed that it may be possible to teach computers to learn everything they need to know about the world and how to carry out tasks for themselves. Machine learning can be understood as an application of AI.

Machine learning was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks. This includes techniques such as Bayesian methods; neural

networks; inductive logic programming; explanation-based, natural language processing; decision tree; and reinforcement learning.

Systems that have hard-coded knowledge bases will typically experience difficulties in new environments. Certain difficulties can be overcome by a system that can acquire its own knowledge. This capability is known as machine learning. This requires knowledge acquisition, inference, updating and refining the knowledge base, acquisition of heuristics, and so forth.

Machine learning is covered in greater depth in Chapter 3.

What Is Data Science?

All AI tasks will use some form of data. Data science is a growing discipline that encompasses anything related to data cleansing, extraction, preparation, and analysis. Data science is a general term for the range of techniques used when trying to extract insights (i.e., trying to understand) and information from data.

The term data science was phrased by William Cleveland in 2001 to describe an academic discipline bringing statistics and computer science closer together.

Teams working on AI projects will undoubtedly be working with data, whether little or big in volume. In the case of big data, real-time data usually demands real-time analytics. In most business cases, a data scientist or data engineer will be required to perform many technical roles related to the data including finding, interpreting, and managing data; ensuring consistency of data; building mathematical models using the data; and presenting and communicating data insights/findings to stakeholders. Although you cannot do big data without data science, you can perform data science without big data—all that is required is data.

Because data exploration requires statistical analysis, many academics see no distinction between data science and statistics. A data science team (even if it is a team of one) is fundamental to deploying a successful project.

The data science team typically performs two key roles. First, beginning with a problem or question, it is seeking to solve it with data; and second, it is using the data to extract insight and intelligence through analytics (Figure 1-3).

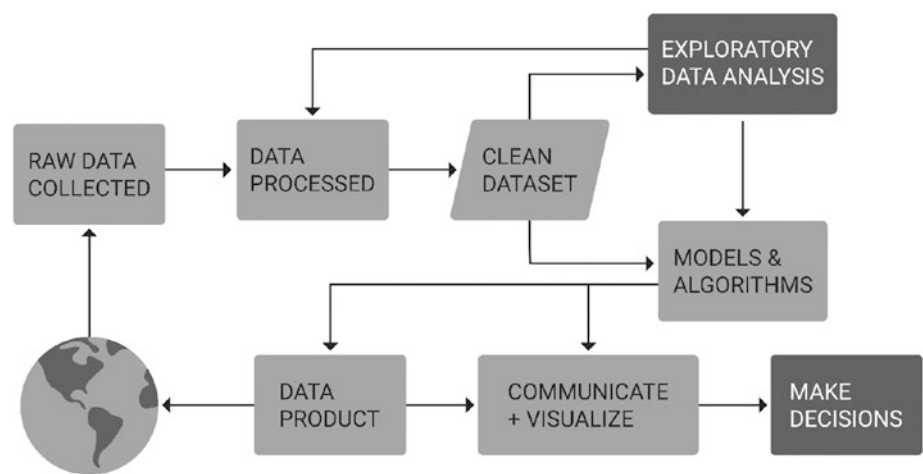


Figure 1-3. *Data science process*

Learning from Real-Time, Big Data

AI had previously been stifled by the technology and data at its disposal. Before the explosion of smartphones and cheaper computing, datasets remained limited in respect to their size, validity, and representative rather than real-time nature.

Today, we have real-time, immediately accessible data and tools that enable rapid analysis. Datafication of modern-day life is fuelling machine learning’s maturity and is facilitating the transition to an evidence-based, data-driven approach. Datafication refers to the modern-day trend of digitalizing (or datafying) every aspect of life). Technology is now agile enough to access these huge datasets to rapidly evolve machine learning applications.

Both patients and healthcare professionals generate a tremendous amount of data. Phones collect metrics such as blood pressure, geographical location; steps walked, nutritional diaries; and other unstructured data such as conversations, reactions, and images.

It's not just digital or clinical data either. Data extraction techniques can be applied to paper documentation, or images scanned, to process the documents for synthesis and recall.

Healthcare professionals collect health biomarkers alongside other structured metrics. Regardless of the source of data, to learn, data must be turned into information. For example, an input of a blood glucose reading from a person with diabetes into an app has far more relevance when blood glucose level targets are known by the system so that it can be understood whether the input met the recommended target range or not.

In the twenty-first century, almost every action leaves some form of transactional footprint. The Apple iPhone 6 is over 32,000 times faster than the Apollo-era NASA computers that took man to the moon. Not only are computers smaller than ever, but they are also more powerful than ever. With organizations capitalizing on sources of big data, there has been a shift toward embedding learnings from data into every aspect of the user experience—from buying a product or service to the user experience within an app.

The value of data is understood when it is taken in its raw form and converted into knowledge that changes practice. This value is driven by project and context. For example, the value may be as the result of faster identification of shortfalls in adherence, compliance, and evidence-based care. It may be better sharing of data and insights within a hospital or organization or more customized relationships with patients to drive adherence, compliance, and boost self-care. Equally, it may be to avoid more costly treatments or costly mistakes.

Applications of AI in Healthcare

It is unlikely artificially intelligent agents will ever completely replace doctors and nurses, but machine learning and AI are transforming the healthcare industry and improving outcomes. Machine learning is improving diagnostics, predicting outcomes, and beginning to scratch the surface of personalized care (Figure 1-4).

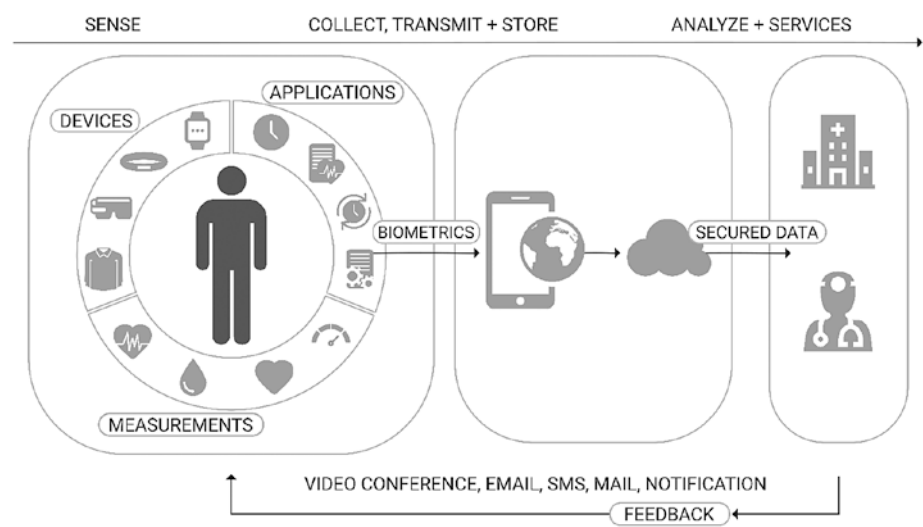


Figure 1-4. A data-driven, patient-healthcare professional relationship

Imagine a situation where you walk in to see your doctor with pain around your heart. After listening to your symptoms, she inputs them into her computer, which pulls up the latest evidence base she should consult to efficiently diagnose and treat your problem. You have an MRI scan, and an intelligent computer system assists the radiologist in detecting any concerns that would be too small for your radiologist’s human eye to see.

Your watch may have even been continuously collecting your blood pressure and pulse while a continuous blood glucose monitor had a

real-time profile of your blood glucose readings. Finally, your medical records and family's medical history is assessed by a computer system that suggests treatment pathways precisely identified to you. Data privacy and governance aside, the implications of what we can learn from combining various pools of data are exciting.

By the end of this book, readers will be able to lead the development of their own machine learning project.

Prediction

Technologies already exist that monitor data to predict disease outbreaks. This is often done using real-time data sources such as social media as well as historical information from the Web and other sources. Malaria outbreaks have been predicted with artificial neural networks, analyzing data including rainfall, temperature, number of cases, and various other data points.

Diagnosis

Many digital technologies offer an alternative to non-emergency health systems. Considering the future, combining the genome with machine learning algorithms provides the opportunity to learn about the risk of disease, improve pharmacogenetics, and provide better treatment pathways for patients.[3]

Personalized Treatment and Behavior Modification

A digital therapy from Diabetes Digital Media, the Low Carb Program, assists people with type 2 diabetes and prediabetes to reverse (i.e., place into remission) their condition. The app provides personalized education

and integrated health tracking, learning from the user's and wider community's progress. At the end of a year, most members who complete the program reduce medication dependency, saving over \$1,015 per patient per year in medication "deprescription." [4]

Drug Discovery

The use of machine learning in preliminary drug discovery has the potential for various uses, from initial screening of drug compounds to predicted success rate based on biological factors. This includes R&D discovery technologies like next-generation sequencing.

Drugs do not necessarily need to be pharmacological in their appearance. The use of digital solutions and aggregation of real-world patient data are providing solutions to conditions once considered to be chronic and progressive. The Low Carb Program app for example, used by over 300,000 people with type 2 diabetes places the condition into remission for 26% of the patients who complete the program at 1-year. [5]

Follow-Up Care

Hospital readmittance is a huge concern in healthcare. Doctors, as well as governments, are struggling to keep patients healthy, particularly when returning home following hospital treatment. Organizations such as NextIT have developed digital health coaches, similar to a virtual customer service representative on an e-commerce site. The assistant prompts questions about the patient's medications and reminds them to take medicine, queries them about their condition symptoms, and conveys relevant information to the doctor.

Realizing the Potential of AI in Healthcare

For AI and machine learning to be fully embraced and integrated within healthcare systems, several key challenges must be addressed.

Understanding Gap

There is a huge disparity between stakeholder understanding and applications of AI and machine learning. Communication of ideas, methodologies, and evaluations are pivotal to the innovation required to progress AI and machine learning in healthcare. Encouraging the adoption of data-driven strategies Data, including the sharing and integration of data, is fundamental to shift healthcare toward realizing precision medicine.

Developing data science teams, focused on learning from data, is key to a successful healthcare strategy. The approach required is one of investment in data. Improving value for both the patient and the provider requires data and hence data science professionals.

Fragmented Data

There are many hurdles to be overcome. Data is currently fragmented and difficult to combine. Patients collect data on their phones, Fitbits, and watches, while physicians collect regular biomarker and demographic data. At no point in the patient experience is this data combined. Nor do infrastructures exist to parse and analyze this larger set of data in a meaningful and robust manner. In addition, electronic health records (EHRs), which at present are still messy and fragmented across databases, require digitizing in a mechanism that is available to patients and providers at their convenience.

Appropriate Security

At the same time, organizations face challenges of security and meeting government regulation specifically with regards to the management of patient data and ensuring its accessibility at all times. What's more, many healthcare institutions are using legacy versions of software that can be more vulnerable to attack. The NHS (National Health Service) digital infrastructure was paralyzed the wrath of the ransomware WannaCry in 2017. The ransomware, which originated in America, scrambled data on computers and demanded payments of \$300 to \$600 to restore access.[6]

Hospitals and GP surgeries in England and Scotland were among over 16 health service organizations hit by the "ransomware" attack. The impact of the attack wasn't just the cost of the technological failure: it had a bearing on patients' lives. Doctors' surgeries and hospitals in some parts of England had to cancel appointments and refuse surgeries. In some areas, people were advised to seek medical care in emergencies only. The NHS was just one of many global organizations crippled through the use of hacking tools; and the ransomware claimed to have infected computers in 150 countries.

Data Governance

With this comes the topic of data governance. Medical data is personal and not easy to access. It is widely assumed that the general public would be reluctant to share their data because of privacy concerns. However, a Wellcome Foundation survey conducted in 2016 on the British public's attitude to commercial access to health data found that 17% of people would never consent to their anonymized data being shared with third parties.[7]

Adhering to multiple sets of regulation means disaster recovery and security is key, and network infrastructure plays a critical role in ensuring these requirements can be met. Healthcare organizations require

modernization of network infrastructure to ensure they are appropriately prepared to provide the best patient care possible. The case for this is highlighted by the fact that the bulk of NHS computers in 2018 use Internet Explorer 8 as their default Internet browser - which was first launched almost a decade ago.[8]

Bias

A significant problem with learning is bias. As AI becomes increasingly interwoven into our daily lives—integrated with our experiences at home, work, and on the road—it is imperative that we question how and why machines do what they do. Within machine learning, learning to learn creates its own inductive bias based on previous experience. Essentially, systems can become biased based on the data environments they are exposed to.

It's not until algorithms are used in live environments that people discover built-in biases, which are often amplified through real-world interactions.

This has expedited the growing need for more transparent algorithms to meet the stringent regulations on drug development and expectation. Transparency is not the only criteria; it is imperative to ensure decision-making is unbiased to fully trust its abilities. People are given confidence through the ability to see through the black box and understand the causal reasoning behind machine conclusions.

Software

Traditional AI systems had been developed in Prolog, Lisp, and ML. Most machine learning systems today are written in Python due to many of the mathematical underpinnings of machine learning that are available as libraries. However, algorithms that “learn” can be developed in most languages, including Perl, C++, Java, and C. This is discussed in more depth in Chapter 3.

Conclusion

The potential applications of machine learning in healthcare are vast and exciting. Intelligent systems can help us reverse disease, detect our risk of cancers, and suggest courses of medication based on real-time biomarkers.

With this also comes tremendous responsibility and questions of wider morality. We don't yet fully understand what can be learned from health data. As a result, the ethics of learning is a fundamental topic for consideration.

On the basis that an intelligent system can detect the risk of disease, should it tell the patient it is tracking the impending outcome? If an algorithm can detect your risk of pancreatic cancer—an often-terminal illness—based on your blood glucose and weight measurements, is it ethical to disclose this to the patient in question? Should such sensitive patient data be shared with healthcare teams - and what the unintended consequences of such actions?

The explosion in digitally connected devices and applications of machine learning are enabling the realization of these once-futuristic concepts, and conversation around these topics is key to progress.

And if a patient opts out of sharing data, is it then ethical to generalize based on known data to predict the same illness risk? And what if I can't get life insurance without this pancreatic cancer check? There are considerable privacy concerns associated with the use of a person's data and what should be private or not, and equally as to what data is useful. Invariably, the more data available, the more precise a decision can be made—but exactly how much is too much is another question. The driving factor is determining the value of data, which will be discussed further in Chapter 2.

The ethics of AI are currently without guidelines, regulations, or parameters on how to govern the enormous treasure chest of data and opportunity. Many assume that AI has an objectivity that puts it above

questions of morality, but that's not the case. AI algorithms are only as fair and unbiased as the learnings, which come from the environmental data. Just like social relationships, political and religious affiliations, and sexual orientation can be revealed from data, learning on health data is revealing new ethical dilemmas for consideration.

Data governance and disclosure of such data still requires policy, at the national and international level. In the future, driverless cars will be able to use tremendous amounts of data in real time to predict the likelihood of survival. Would it be ethical for the systems to choose who lives or dies or for a doctor to decide whom to treat based on the reading from two patients' Apple Watches?

In reality, this is just the beginning. As technologies develop, new and improved treatments and diagnoses will save more lives and cure more diseases. With the future of medicine grounded in data and analytics, it begs the question as to whether there will ever be enough data.