INTRODUCTION

In the pages of this book you will read about the myth of artificial intelligence. The myth is not that true AI is possible. As to that, the future of AI is a scientific unknown. The myth of artificial intelligence is that its arrival is inevitable, and only a matter of time—that we have already embarked on the path that will lead to human-level AI, and then superintelligence. We have not. The path exists only in our imaginations. Yet the inevitability of AI is so ingrained in popular discussion—promoted by media pundits, thought leaders like Elon Musk, and even many AI scientists (though certainly not all)—that arguing against it is often taken as a form of Luddism, or at the very least a shortsighted view of the future of technology and a dangerous failure to prepare for a world of intelligent machines.

As I will show, the science of AI has uncovered a very large mystery at the heart of intelligence, which no one currently has a clue how to solve. Proponents of AI have huge incentives to minimize its known limitations. After all, AI is big business, and it's increasingly dominant in culture. Yet the possibilities for future AI systems are limited by what we currently know about the nature of intelligence, whether we like it or not. And here we should say it directly: all evidence suggests that human and machine intelligence are radically different. The myth of AI insists that the differences are only temporary, and that more powerful systems will eventually erase them. Futurists like

Ray Kurzweil and philosopher Nick Bostrom, prominent purveyors of the myth, talk not only as if human-level AI were inevitable, but as if, soon after its arrival, superintelligent machines would leave us far behind.

This book explains two important aspects of the AI myth, one scientific and one cultural. The scientific part of the myth assumes that we need only keep "chipping away" at the challenge of general intelligence by making progress on narrow feats of intelligence, like playing games or recognizing images. This is a profound mistake: success on narrow applications gets us not one step closer to general intelligence. The inferences that systems require for general intelligence—to read a newspaper, or hold a basic conversation, or become a helpmeet like Rosie the Robot in *The Jetsons*—cannot be programmed, learned, or engineered with our current knowledge of AI. As we successfully apply simpler, narrow versions of intelligence that benefit from faster computers and lots of data, we are not making incremental progress, but rather picking low-hanging fruit. The jump to general "common sense" is completely different, and there's no known path from the one to the other. No algorithm exists for general intelligence. And we have good reason to be skeptical that such an algorithm will emerge through further efforts on deep learning systems or any other approach popular today. Much more likely, it will require a major scientific breakthrough, and no one currently has the slightest idea what such a breakthrough would even look like, let alone the details of getting to it.

Mythology about AI is bad, then, because it covers up a scientific mystery in endless talk of ongoing progress. The myth props up belief in inevitable success, but genuine respect for science should bring us back to the drawing board. This brings us to the second subject of these pages: the cultural consequences of the myth. Pursuing the myth is not a good way to follow "the smart money," or even a neutral stance. It is bad for science, and it is bad for us. Why? One reason is

that we are unlikely to get innovation if we choose to ignore a core mystery rather than face up to it. A healthy culture for innovation emphasizes exploring unknowns, not hyping extensions of existing methods—especially when these methods have been shown to be inadequate to take us much further. Mythology about inevitable success in AI tends to extinguish the very culture of invention necessary for real progress—with or without human-level AI. The myth also encourages resignation to the creep of a machine-land, where genuine invention is sidelined in favor of futuristic talk advocating current approaches, often from entrenched interests.

Who should read this book? Certainly, anyone should who is excited about AI but wonders why it is always ten or twenty years away. There is a scientific reason for this, which I explain. You should also read this book if you think AI's advance toward superintelligence is inevitable and worry about what to do when it arrives. While I cannot prove that AI overlords will not one day appear, I can give you reason to seriously discount the prospects of that scenario. Most generally, you should read this book if you are simply curious yet confused about the widespread hype surrounding AI in our society. I will explain the origins of the myth of AI, what we know and don't know about the prospects of actually achieving human-level AI, and why we need to better appreciate the only true intelligence we know—our own.

IN THIS BOOK

In Part One, The Simplified World, I explain how our AI culture has simplified ideas about people, while expanding ideas about technology. This began with AI's founder, Alan Turing, and involved understandable but unfortunate simplifications I call "intelligence errors." Initial errors were magnified into an ideology by Turing's friend and statistician, I. J. Good, who introduced the idea of "ultraintelligence" as the predictable result once human-level AI had been achieved.

Between Turing and Good, we see the modern myth of AI take shape. Its development has landed us in an era of what I call technological kitsch—cheap imitations of deeper ideas that cut off intelligent engagement and weaken our culture. Kitsch tells us how to think and how to feel. The purveyors of kitsch benefit, while the consumers of kitsch experience a loss. They—we—end up in a shallow world.

In Part Two, The Problem of Inference, I argue that the only type of inference—thinking, in other words—that will work for human-level AI (or anything even close to it) is the one we don't have a clue how to program or engineer. The problem of inference goes to the heart of the AI debate because it deals directly with intelligence, in people or machines. Our knowledge of the various types of inference dates back to Aristotle and other ancient Greeks, and has been developed in the fields of logic and mathematics. Inference is already described using formal, symbolic systems like computer programs, so a very clear view of the project of engineering intelligence can be gained by exploring inference. There are three types. Classic AI explored one (deduction), modern AI explores another (induction). The third type (abduction) makes for general intelligence, and, surprise, no one is working on it—at all.[1] Finally, since each type of inference is distinct—meaning, one type cannot be reduced to another—we know that failure to build AI systems using the type of inference undergirding general intelligence will result in failure to make progress toward artificial general intelligence, or AGI.

In Part Three, The Future of the Myth, I argue that the myth has very bad consequences if taken seriously, because it subverts science. In particular, it erodes a culture of human intelligence and invention, which is necessary for the very breakthroughs we will need to understand our own future. Data science (the application of AI to "big data") is at best a prosthetic for human ingenuity, which if used correctly can help us deal with our modern "data deluge." If used as a replacement for individual intelligence, it tends to chew up investment without delivering results. I explain, in particular, how the myth has negatively affected research in neuroscience, among other recent scientific pursuits. The price we are paying for the myth is too high. Since we have no good scientific reason to believe the myth is true, and every reason to reject it for the purpose of our own future flourishing, we need to radically rethink the discussion about AI.

Chapter 1

. . .

THE INTELLIGENCE ERROR

The story of artificial intelligence starts with the ideas of someone who had immense human intelligence: the computer pioneer Alan Turing.

In 1950 Turing published a provocative paper, "Computing Machinery and Intelligence," about the possibility of intelligent machines.[1] The paper was bold, coming at a time when computers were new and unimpressive by today's standards. Slow, heavy pieces of hardware sped up scientific calculations like code breaking. After much preparation, they could be fed physical equations and initial conditions and crank out the radius of a nuclear blast. IBM quickly grasped their potential for replacing humans doing calculations for businesses, like updating spreadsheets. But viewing computers as "thinking" took imagination.

Turing's proposal was based on a popular entertainment called the "imitation game." In the original game, a man and a woman are hidden from view. A third person, the interrogator, relays questions to one of them at a time and, by reading the answers, attempts to determine which is the man and which the woman. The twist is that the man has to try to deceive the interrogator while the woman tries to assist him—making replies from either side suspect. Turing replaced the man and woman with a computer and a human. Thus began what we now call the Turing test: a computer and a human receive typed questions from a human judge, and if the judge can't accurately identify which is the computer, the computer wins. Turing argued that with such an outcome, we have no good reason to define the machine as unintelligent, regardless of whether it is human or not. Thus, the question of whether a machine has intelligence replaces the question of whether it can truly think.

The Turing test is actually very difficult—no computer has ever passed it. Turing, of course, didn't know this long-term result in 1950; however, by replacing pesky philosophical questions about "consciousness" and "thinking" with a test of observable output, he encouraged the view of AI as a legitimate science with a well-defined aim. As AI took shape in the 1950s, many of its pioneers and supporters agreed with Turing: any computer holding a sustained and convincing conversation with a person would be, most of us would grant, doing something that requires thinking (whatever that is).

TURING'S INTUITION / INGENUITY DISTINCTION

Turing had made his reputation as a mathematician long before he began writing about AI. In 1936, he published a short mathematical paper on the precise meaning of "computer," which at the time referred to a person working through a sequence of steps to get a definite result (like performing a calculation).[2] In this paper, he replaced the human computer with the idea of a machine doing the same work. The paper ventured into difficult mathematics. But in its treatment of machines it made no reference to human thinking or the mind. Machines can run automatically, Turing said, and the problems they solve do not require any "external" help, or intelligence. This external intelligence—the human factor—is what mathematicians sometimes call "intuition."

Turing's 1936 work on computing machines helped launch computer science as a discipline and was an important contribution to mathematical logic. Still, Turing apparently thought that his early definition missed something essential. In fact, the same idea of the mind or human faculties assisting problem-solving appeared two years later in his PhD thesis, a clever but ultimately unsuccessful attempt to bypass a result from the Austrian-born mathematical logician Kurt Gödel (more on this later). Turing's thesis contains this curious passage about intuition, which he compares with another mental capability he calls ingenuity:

> Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two faculties, which we may call intuition and ingenuity. The activity of the intuition consists in making spontaneous judgments which are not the result of conscious trains of reasoning. These judgments are often but by no means invariably correct (leaving aside the question as to what is meant by "correct"). Often it is possible to find some other way of verifying the correctness of an intuitive judgment. One may for instance judge that all positive integers are uniquely factorable into primes; a detailed mathematical argument leads to the same result. It will also involve intuitive judgments, but they will be ones less open to criticism than the original judgment about factorization. I shall not attempt to explain this idea of "intuition" any more explicitly.

Turing then moves on to explain ingenuity: "The exercise of ingenuity in mathematics consists in aiding the intuition through suitable arrangements of propositions, and perhaps geometrical figures or drawings. It is intended that when these are really well arranged the validity of the intuitive steps which are required cannot seriously be doubted."[3]

Though his language is framed for specialists, Turing is pointing out the obvious: mathematicians typically select problems or "see" an interesting problem to work on using some capacity that at least *seems* indivisible into steps—and therefore not obviously amenable to computer programming.

GÖDEL'S INSIGHT

Gödel, too, was thinking about mechanical intelligence. Like Turing, he was obsessed with the distinction between ingenuity (mechanics) and intuition (mind). His distinction was essentially the same as Turing's, in different language: proof versus truth (or "proof-theory" versus "model-theory" in mathematics lingo). Are the concepts of proof and truth, Gödel wondered, in the end the same? If so, mathematics and even science itself might be understood purely mechanically. Human thinking in this view would be mechanical, too. The concept of AI, though the term remained to be coined, hovered above the question. Is the mind's intuition, its ability to grasp truth and meaning, reducible to a machine, to computation?

This was Gödel's question. In answering it, he ran into a snag that would soon make him world-famous. In 1931, Gödel published two theorems of mathematical logic known as his incompleteness theorems. In them, he demonstrated the inherent limitations of all formal mathematical systems. It was a brilliant stroke. Gödel showed unmistakably that mathematics—*all* of mathematics, with certain straightforward assumptions—is, strictly speaking, not mechanical or "formalizable." More specifically, Gödel proved that there must exist some statements in any formal (mathematical or computational) system that are True, with capital-T standing, yet not provable in the system itself using any of its rules. The True statement can be recognized by a human mind, but is (provably) not provable by the system it's formulated in.

How did Gödel reach this conclusion? The details are complicated and technical, but Gödel's basic idea is that we can treat a mathematical system complicated enough to do addition as a system of meaning, almost like a natural language such as English or German—and the same applies to all more complicated systems. By treating it this way, we enable the system to talk about itself. It can say about itself, for instance, that it has certain limitations. This was Gödel's insight.

Formal systems like those in mathematics allow for the precise expression of truth and falsehood. Typically, we establish truth by using the tools of proof—we use rules to prove something, so we know it's definitely true. But are there true statements that can't be proven? Can the mind know things the system cannot? In the simple case of arithmetic, we express truths by writing equations like "$2 + 2 = 4$." Ordinary equations are true statements in the system of arithmetic, and they are provable using the rules of arithmetic. Here, provable equals true. Mathematicians before Gödel thought all of mathematics had this property. This implied that machines could crank out all truths in different mathematical systems by simply applying the rules correctly. It's a beautiful idea. It's just not true.

Gödel hit upon the rare but powerful property of self-reference. Mathematical versions of self-referring expressions, such as "This statement is not provable in this system," can be constructed without breaking the rules of mathematical systems. But the so-called self-referring "Gödel statements" introduce contradictions into mathematics: if they are true, then they are unprovable. If they are false, then because they say they are unprovable, they are actually true. True means false, and false means true—a contradiction.

Going back to the concept of intuition, we humans can see that the Gödel statement is in fact true, but because of Gödel's result, we also know that the system's rules can't prove it—the system is in effect blind to something not covered by its rules.[4] Truth and provability pull apart. Perhaps mind and machine do, as well. The purely formal system has limits, at any rate. It cannot prove in its own language something that is true. In other words, we can see something that the computer cannot.[5]

Gödel's result dealt a massive blow to a popular idea at the time, that all of mathematics could be converted into rule-based operations, cranking out mathematical truths one by one. The zeitgeist was formalism—not talk of minds, spirits, souls, and the like. The formalist movement in mathematics signaled a broader turn by intellectuals toward scientific materialism, and in particular, logical positivism—a movement dedicated to eradicating traditional metaphysics like Platonism, with its abstract Forms that couldn't be observed with the senses, and traditional notions in religion like the existence of God. The world was turning to the idea of precision machines, in effect. And no one took up the formalist cause as vigorously as the German mathematician David Hilbert.

HILBERT'S CHALLENGE

At the outset of the twentieth century (before Gödel), David Hilbert had issued a challenge to the mathematical world: show that all of mathematics rested on a secure foundation. Hilbert's worry was understandable. If the purely formal rules of mathematics can't prove any and all truths, it's at least theoretically possible for mathematics to disguise contradictions and nonsense. A contradiction buried somewhere in mathematics ruins everything, because from a contradiction anything can be proven. Formalism then becomes useless.

Hilbert expressed the dream of all formalists, to prove finally that mathematics is a closed system governed only by rules. Truth is just "proof." We acquire knowledge by simply tracing the "code" of a proof and confirming no rules were violated. The larger dream, thinly disguised, was really a worldview, a picture of the universe as itself a mechanism. AI began taking shape as an idea, a philosophical position that might also be proven. Formalism treated intelligence as a rule-based process. A machine.

Hilbert issued his challenge at the Second International Congress of Mathematicians in Paris in 1900. The intellectual world was listening. His challenge had three main parts: to prove that mathematics was complete; to prove that mathematics was consistent; and to prove that mathematics was decidable.

Gödel dealt the first and second parts of Hilbert's challenge a death blow with the publication of his incompleteness theorems in 1931. The question of decidability was left unanswered. A system is decidable if there is a definite procedure (a proof, or sequence of deterministic, obvious steps) to establish whether any statement constructed using the rules of the system is true or false. The statement $2+2=4$ must be True, and $2+2=5$ must be False. And so for all statements that one can validly make using the symbols and rules of the system. Since arithmetic was thought to be the foundation of mathematics, proving mathematics was decidable amounted to proving the result for arithmetic and its extensions. This would amount to saying that mathematicians, playing a "game" with rules and symbols (the formalist idea), were in fact playing a valid game that never led to contradiction or absurdity.

Turing was fascinated with Gödel's result, which demonstrated not the power of formal systems but rather their limitations. He took up work on the remaining part of Hilbert's challenge, and began thinking in earnest about whether a decision procedure for formal systems might exist. By 1936, in his paper "Computable Numbers," he proved that it must not. Turing realized that Gödel's use of self-reference also applied to questions about decision procedures or, in effect, computer programs. In particular, he realized that there must exist (real) numbers that *no* definite method could "calculate," by writing out their decimal expansion, digit by digit. He imported a result from the nineteenth-century mathematician Georg Cantor, who proved that real numbers (those with a decimal expansion) were more numerous than the integers, even though real numbers and integers are both infinite. Turing stood on the shoulders of giants, perhaps. But in the end, his work in "Computable Numbers" proved again a negative. It was a limiting result: no universal decision procedure was possible. In other words, rules—even in mathematics—aren't enough. Hilbert was wrong.[6]

IMPLICATIONS FOR AI

What is important to AI here is this: Turing disproved that mathematics was decidable by inventing a machine, a deterministic machine, requiring no insight or intelligence to solve problems. Today, we refer to his abstract formulation of a machine as a Turing machine. I am typing on one right now. Turing machines are computers. It is one of the great ironies of intellectual history that the theoretical framework for computation was put in place as a side-thought, a means to another end. While working to disprove that mathematics itself was decidable, Turing first invented something precise and mechanical, the computer.

In his 1938 PhD thesis, Turing hoped that formal systems might be extended by including additional rules (then sets of rules, and sets of sets of rules) that could handle the "Gödel problem." He discovered, rather, that the new, more powerful system would have a new, more complicated Gödel problem. There was no way around Gödel's incompleteness. Buried in the complexities of Turing's discussion of formal systems, however, is an odd suggestion, relevant to the possibility of AI. Perhaps the faculty of intuition cannot be reduced to an algorithm, to the rules of a system?

Turing wanted to find a way out of Gödel's limiting result in his 1938 thesis, but he discovered that this was impossible. Instead, he switched gears, exploring how, as he put it, to "greatly reduce" the requirement of human intuition when doing calculations. His thesis considered the powers of ingenuity, by creating ever more complicated systems of rules. (Ingenuity, it turned out, could become universal—there are machines that can take as input other machines, and thus run all the machines that can be built. This insight, technically a universal Turing machine and not a simple Turing machine, was to become the digital computer.) But in his formal work on computing, Turing had (perhaps inadvertently) let the cat out of the bag. By allowing for intuition as distinct from and outside of the operations of a purely formal system like a computer, Turing in effect suggested that there may be differences between computer programs that do math and mathematicians.

It was a curious turn, therefore, that Turing made from his early work in the 1930s to the more wide-ranging speculation about the possibility of intelligent computers in "Computing Machinery and Intelligence," published a little over a decade later. By 1950, discussion of intuition disappeared from Turing's writings about the implications of Gödel. His interests turned, in effect, to the possibility that computers might become "intuition-machines" themselves. In essence, he decided that Gödel's result didn't apply to the question of AI: if we humans are highly advanced computers, Gödel's result means only that there are some statements that we cannot understand or see to be true, just as with less complicated computers. The statements might be fantastically complicated and interesting. Or, possibly, they might be banal yet overwhelmingly complex. Gödel's result left open the question of whether minds were just very complicated machines, with very complicated limitations.

Intuition, in other words, had become part of Turing's ideas about machines and their powers. Gödel's result couldn't say (to Turing, anyway) whether minds were machines or not. On the one hand, incompleteness says that some statements can be seen to be true using intuition, but cannot be proved by a computer using ingenuity. On the other hand, a more powerful computer can use more axioms (or more bits of relevant code) and prove the result—thus showing that intuition is not beyond computation for that problem. This becomes an arms race: more and more powerful ingenuity substituting for intuition on more and more complicated problems. No one can say who wins the race, so no one can make a case—using the incompleteness result—about the inherent differences between intuition (mind) and ingenuity (machine). But as Turing no doubt knew, if this were true, then so too was at least the possibility of artificial intelligence.

Thus, between 1938 and 1950, Turing had a change of heart about ingenuity and intuition. In 1938, intuition was the mysterious "power of selection" that helped mathematicians decide which systems to work with and what problems to solve. Intuition was not something in the computer. It was something that decided things about the computer. In 1938, Turing thought intuition wasn't part of any system, which suggested not only that minds and machines were fundamentally different but that AI-as-human-thinking was well-nigh impossible.

Yet by 1950 he had reversed his position. With the Turing test, he offered a challenge for skeptics and a sort of defense of intuition in machines, asking in effect: Why not? This was a radical about-face. A new view of intelligence, it seemed, was taking shape.

Why the shift? Something outside the world of strict mathematics and logic and formal systems had happened to Turing between 1938 and 1950. It had happened, in fact, to all of Great Britain, and indeed to most of the world. What happened was the Second World War.

Chapter 2

. . .

TURING AT BLETCHLEY

The game of chess fascinated Turing—as it did his wartime colleague, mathematician I. J. "Jack" Good. The two would play against each other (Good usually won) and work out decision procedures and rules of thumb for winning moves. Playing chess involves following the rules of the game (ingenuity), and it also seems to require insight (intuition) into which rules to choose given different positions on the game board. To win at chess, it is not enough to apply the rules; you have to know which rules to select in the first place.

Turing saw chess as a handy (and no doubt entertaining) way to think about machines and the possibility of giving them intuition. Across the Atlantic, the founder of modern information theory, Turing's colleague and friend Claude Shannon at Bell Labs, was also thinking about chess. He later built one of the first chess-playing computers, an extension of work he had done earlier on a proto-computer called the "differential analyzer," which could convert certain problems in calculus into mechanical procedures.[1]

THE SIMPLIFICATION OF INTELLIGENCE BEGINS

Chess fascinated Turing and his colleagues in part because it seemed that a computer could be programmed to play it, without the human programmer needing to know everything in advance. Because