1 Introduction

Prediction, Understanding, and Deep Neural Networks

The July 2008 issue of *WIRED* magazine is an unusual presence in the archives of the influential technology publication. That year was an exciting time in tech: The iPhone was only a year old, and the possibilities of the smartphone age were just coming into view. One article from the issue details the development of the then-upcoming Android operating system; another speculates, presciently, that mobile devices might represent the future of gaming.

But twenty-first-century consumer technologies play second fiddle on this issue's cover, which features outdated instrumentation ranging from a bulky, mid-century camera to an antique-styled telescope. The theme, declared in big white letters, is "The End of Science," and in the cover story, editor-in-chief Chris Anderson consigned more than just these old tools to the dustbin of history. Scientific theory itself, he declared, was over. This was the "Petabyte Age," in which large-scale datasets would entirely obviate the need for scientific insight.

"Petabytes allow us to say: 'Correlation is enough,'" wrote Anderson. "We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

clusters the world has ever seen and let statistical algorithms find patterns where science cannot."[1]

Anderson wasn't wrong to identify a shift in the scale of scientific data. At the time he was writing, the Large Hadron Collider (LHC) was about to come online; it would eventually produce multiple petabytes of usable data each day. But the idea that "Big Data" would herald the end of scientific theory found a decidedly cool reception among those who worked with data and theories on a daily basis. This being 2008, many scientists took to their personal blogs to criticize Anderson's "hubris"[2] and "wrong-headed[ness]."[3] Some reactions made it into print: In the journal *EMBO Reports*, the philosopher Massimo Pigliucci wrote that Anderson "doesn't understand much about either science or the scientific method."[4]

Time has borne out many of those criticisms: Seventeen years on, science has not yet collapsed into a morass of correlations. Scientists still propose and refine new theories, use models to analyze data, and work to identify causes for the phenomena they observe. Trying to understand how the universe works hasn't yet gone out of fashion.

And yet, Anderson is starting to look less like a crackpot and more like a Cassandra. In 2023, the academic journal *Nature* polled 1,600 scientists and found that 69 percent of them worried that new technologies could "lead to more reliance on pattern recognition without understanding" – precisely the scenario that Anderson anticipated, albeit without his maverick optimism.[5] The reason for their concern, however, wasn't the "petabyte era" of which Anderson had made so much hay. It was artificial intelligence.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

Back when Anderson wrote his article, scientists still lacked the tools to take full advantage of their new large-scale data resources. Positing an end to scientific theory was like declaring a gold rush before the invention of panning. But things have changed. With the advent of modern AI techniques, especially the brain-inspired tools known as deep neural networks, scientists can now squeeze a huge amount of utility from datasets far too big to be analyzed by hand. These AI systems can identify subtle patterns in data that exceed the boundaries of human comprehension, then use those patterns to make reliable, useful predictions.

In disciplines as diverse as neuroscience, biochemistry, and climate science, the marriage of big data and AI has yielded extraordinary practical advancements. Speech-restoring brain implants, new snake antivenoms, and cheaper and more accurate weather prediction will meaningfully improve human lives. Such accomplishments have been widely, and rightfully, lauded: The 2024 Nobel prizes in both physics and chemistry were awarded for achievements related to neural networks.

But deep neural networks specifically, and modern AI more generally, are not panaceas. The late philosopher Paul Humphreys, who studied computer-based scientific methods, proposed a distinction between the "practical view" and the "intellectual view" of science.[6] According to the practical view, the aim of science is to devise useful tools: instruments, software, and strategies that allow people to predict and manipulate the world around them. Under the intellectual view, however, science is simply about figuring out how the universe works, independent of whether

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

that understanding leads to technological advancement (though it often does).

Scientists make all sorts of different predictions in the service of practical goals. Some, in alignment with the typical English meaning of "prediction," concern future events, as when a meteorologist uses a weather model to predict a hurricane. Others pertain to phenomena that already exist but that we don't yet know about in detail, as when a doctor infers that a patient has a disease based on the results of a blood test. In both cases, intellectual goals are a key step on the way to practical ones: Most weather models require detailed physics equations, and the interpretation of blood tests is based on our current understanding of human physiology and disease processes. In order to make reliable and robust predictions about some part of the universe, or devise tools that could allow others to do the same, you need to have a solid mental grasp of how that part of the universe works. In much of science, intellectual and practical goals are near impossible to disentangle.

AI, however, is changing all of that. The deep neural networks that are enabling astonishing practical advances across the sciences offer few clues about the nature of the data they were trained on or the intellectual basis of their predictions. The very best scientific AI tools perform end runs around human insight while making far more accurate and useful predictions than our quaint theories enable. That's nothing short of revolutionary.

This book is about the prediction revolution that AI has enabled and what that revolution means for the present and future of science. In the chapters that follow, I will detail

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

how dozens of scientists, from particle physicists to neuroscientists, are using AI and thinking about its role in their work. But before we consider how AI is changing the way scientists make predictions, we need to get a clear picture of where predictions used to come from.

The Prediction That Launched a Voyage

In August of 1768, Lieutenant James Cook departed from England aboard the ship *Endeavour* with a crew of ninety-four men. Their destination was Tahiti, on the far side of the world; though Cook would later achieve considerable fame for his three journeys to the Pacific, at this point in his life he had never traveled farther west than Quebec. To make matters more challenging, Cook was on a strict timeline. He absolutely needed to reach Tahiti, and construct an astronomical observatory there, before June 3 of the following year, because that was the day on which astronomers had predicted that Venus's shadow would pass over the surface of the Sun. By collecting data about this "transit of Venus" from locations all over the globe and integrating those observations using clever geometry, astronomers would be able to calculate the distance between the Earth and the Sun.

Traveling such distances in the eighteenth century wasn't cheap. The British government spent around £13,000 to fit out Cook's ship and pay his men – about £2,000,000 today, or 2,700,000 USD. Though observing the transit of Venus was not the only goal of Cook's voyage, the Crown's enormous commitment of funds testifies to its confidence that the astronomical event would occur on its predicted

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

date. That confidence was attributable to the Herculean scientific efforts of two men – one who specialized in collecting data, and one who excelled at taming it.

Over a century and a half earlier, the astronomer Tycho Brahe had devoted decades of his life to recording extensive celestial observations – first on an island near Copenhagen, and then in the city of Prague. A clever instrumentalist and meticulous documentarian, Tycho (as he's typically known) amassed one of the first collections that could be called "big data." In an early twentieth-century edition of Tycho's complete works, his extant astronomical observations total over 1,500 dense pages.[7]

After Tycho's death in 1601, it fell to his assistant, Johannes Kepler, to make sense of all of the data. Unlike Tycho, who espoused a bespoke theory of the solar system in which the Sun orbited the Earth and the other planets orbited the Sun, Kepler was an adherent of Copernicus's heliocentrism. From that vantage, he analyzed Tycho's observations of Mars and, after years of arduous calculation, convinced himself that the planet's orbit around the Sun could not possibly be circular. Mars and its fellow planets, he determined, trace out slightly oblong or elliptical paths, with the Sun not quite at the center. This was Kepler's first law of planetary motion. He extracted two more laws from Tycho's data: his second law, which stipulates that an imaginary line drawn between a planet and the Sun will sweep out an equal area in an equal interval of time, such that planets travel faster when they are nearer the Sun; and his third, which describes the precise mathematical relationship between a planet's distance from the Sun and the time it takes to complete a full orbit.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

These three laws allowed Kepler to predict the future positions of the planets with unprecedented accuracy. That ability would end up being a burden: For the final decade of his life, Kepler was obligated to devote much of his time to onerous computations. He spent years preparing books of tables, called ephemerides, from which astronomers, sailors, or other interested parties could calculate or read off the positions of the planets on a specific date. To one eager reader of his work he wrote, "I ask you, my friends, not to condemn me entirely to the mill of mathematical calculations, but to permit me time for philosophical speculations, my only delights."[8]

He did find some enjoyment among all that math. While tabulating his ephemeris for the year 1631, Kepler stumbled upon what he called a "rare and wonderful phenomenon": On December 6, the shadow of Venus would cross over the surface of the Sun.[9] With his son-in-law, Kepler published and circulated a small pamphlet urging astronomers across Europe to observe this event, in order to determine the relative size of Venus compared with the Sun. He drummed up a sense of urgency among his readers by emphasizing that another transit of Venus would not occur for 130 years. Some astronomers listened to his admonition, but in vain – the transit did indeed occur, but it was not visible from western Europe. The astronomical community settled in for a long wait.

But Kepler had made an error. In October of 1639, a twenty-year-old Kepler devotee named Jeremiah Horrocks redid Kepler's Venus calculations and discovered, to his surprise, that the planet was due to transit across the Sun in only

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

a month. On that day, November 24, Horrocks set up a telescope indoors to project the Sun's light onto a sheet of paper and then waited anxiously to see if the clouds would clear in time for the 3 p.m. transit. They did. Horrocks, and a friend elsewhere in England whom he had managed to notify, were the first humans ever to record a transit of Venus. In doing so, they proved the predictive precision of Kepler's laws.

Half a century later, the famed astronomer Edmond Halley leveraged Kepler's calculations and insights to predict the next two transits of Venus, in 1761 and 1769. Like Kepler, he urged that these transits be observed, but with a different motivation: He had realized that, if a transit of Venus were observed from multiple distant points on the globe, the difference in duration of the transit from location to location could be used to infer the distance between the Earth and the Sun. He published his argument in 1716 and so gave would-be observers plenty of time to prepare.[10] For the 1761 transit, the British and French governments sent out several major expeditions, but for the most part they were thwarted by trouble at sea or poor weather on the day.

Thankfully, the astronomical community soon had another shot. They made the most of it. James Cook arrived to Tahiti in April of 1769 for the anticipated June transit, and his men completed the astronomical observatory in early May, over a month before it would be needed.

The long predicted day dawned serendipitously – the impeccably clear sky was, in Cook's words, "as favourable to our purpose as we could wish."[11] At 7:21 in the morning Tahiti time, as anticipated, a small sliver of black appeared at the edge of the Sun, and over the next fifteen minutes the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

sliver grew into a disk – the shadow of Venus. For the next several hours, in the Tahitian heat, Cook and his men peered through telescopes and watched Venus's shadow slowly crawl across the Sun's surface until it reached the opposite edge, just after 1 p.m. Based off of their observations, as well as measurements taken in Russia, Denmark, modern-day Mexico, and modern-day Canada, the Oxford astronomer Thomas Hornsby calculated the distance between the Earth and the Sun to be 93,726,900 miles – astonishingly close to the currently accepted value of 92,955,800 miles.[12]

That calculation, and the transoceanic journeys that enabled it, ultimately owed their existence to Kepler, whose laws made it possible to predict astronomical events with precision. But characterizing those laws as predictive tools sells their virtues short. Prediction can be cheap: Drop a copy of Kepler's 1630 *Ephemerides* in the lap of any literate seventeenth-century person – perhaps a pastor, or a university student – and he could tell you where in the sky to seek out the celestial body of your choosing, on any day in the next six years. If prediction is all you're after, the pastor might be just as useful as Kepler himself.

Unlike Kepler, however, the pastor would have no sense of how those endless columns of numbers all hung together. His predictions might be top-notch, but he would still lack Kepler's understanding.

The Shape of Scientific Understanding

It's worth taking a moment to get precise about what it means to say that Kepler "understood" Tycho's Mars data,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

or his own *Ephemerides*, or the solar system more generally. Over the past couple of decades, a rich philosophical discussion about understanding and its role in the scientific project has developed, and it would take an entire additional book to do justice to that discourse. But you can develop a general sense of the term by thinking about a conversation you might have with a young relative. When they ask "Why do dogs have four legs?" or "How does a car work?," you answer those questions (if you are able to) by drawing on your understanding – your knowledge and insight – about evolutionary biology or combustion engines. The richer your understanding of a given topic, the more how and why questions you can answer about it. And when you answer those questions for your young relative, you pass your understanding on to them – or, at least, you try to.

Someone understands a phenomenon if they both have an explanation for that phenomenon and know how to apply that explanation. In science, explanations – answers to those why and how questions – take on a variety of different forms. They might draw on laws, like Kepler's; models, like Copernicus's heliocentric picture of the solar system; or specific causes, like the swat from a mischievous cat that sends an inkwell tumbling to the floor. But having an explanation isn't enough to achieve understanding. Anyone can parrot Kepler's laws in response to questions about the locations of planets in the sky, even if they have no idea what an ellipse is. For your knowledge of Kepler's laws to count in favor of your understanding of the solar system, you have to do more than recite the words. You have to know how to use them.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

Kepler's success in producing his ephemerides, and his accurate prediction of the 1631 transit of Venus, demonstrate that he knew how to apply his laws. He could also have used them to predict the trajectories of the planets in what philosophers might call counterfactual universes, or universes different from the one in which we live – such as one where Mars sits somewhere between the Earth and Venus. (According to Kepler's third law, in such a universe, a Martian year would last around three hundred Earth days.)

This ability to reason through counterfactuals – something that likely would have been beyond the reach of the literate pastor – is characteristic of understanding. It also has practical applications, despite its heady vibe: Counterfactual reasoning is what allows scientists to anticipate the effects that our actions might have on the world. Kepler himself could hardly have envisioned changing the trajectory of one of the heavenly bodies; indeed, such an idea might have struck him as irreligious. But in 2022, NASA scientists intentionally flew a spacecraft directly into the asteroid Dimorphos and shifted its orbit around its larger, partner asteroid, Didymos, in order to demonstrate that such a tactic could effectively deflect an asteroid on a collision course with Earth. It's not inconceivable, then, that Kepler's laws, and the counterfactual predictions they enable, could one day help astronomers save the planet.

Kepler's counterfactual skills indicate that he understood the motion of the planets well – better than any human ever had at the time when he was writing. But his abilities weren't boundless. If you asked him what would happen to Mars's orbit if the mass of the Sun halved or doubled, he

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

would be flummoxed. Less exotic questions could stump him as well. Ask why Mars sweeps out its particular path across the sky, and he would offer a thorough answer grounded in the ellipticity of its orbit – but press further and ask why Mars follows an elliptical orbit, and he would have no response.

No one could offer a satisfactory answer to that question until about thirty-five years after Kepler's death. In 1665, Cambridge University shut down due to an outbreak of the bubonic plague and sent home all of its students, including Isaac Newton. Stuck at home in the rural town of Woolsthorpe, Newton passed his time by laying the groundwork for all of modern physics. In just two years, he developed a sophisticated theory of color by experimenting with glass prisms, established the foundations of calculus, and realized that the Moon was held in orbit around the Earth by the force of gravity – famously, in this last case, upon observing the fall of an apple from a tree. Over the next two decades, he refined his theory, eventually discovering that all objects are attracted to each other by a force proportional to the product of their masses divided by the square of the distance between them – his law of universal gravitation. In his 1687 *Philosophiæ naturalis principia mathematica*, where he first announced this law, Newton explicitly demonstrates that all three of Kepler's laws follow from his insights.

Newton and Kepler show us that understanding comes in degrees. Though Kepler certainly understood the motions of the planets, it's indisputable that Newton understood them better. Often, deeper understanding comes from the ability to explain explanations: The motion of Mars is explained by Kepler's laws, which, in turn, are explained by

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

Newton's law of universal gravitation. And the chain doesn't end there. Einstein's theory of general relativity explains the gravitational force that Newton described as a consequence of the curving of spacetime by massive objects, and perhaps some future "theory of everything" will one day explain general relativity.

Superior understanding also grants greater predictive powers. The law of universal gravitation makes it possible to account for the gravitational effects the planets have on one another, so Newton enabled astronomers to make more accurate predictions about the positions of the planets than were possible using Kepler's laws alone. And better predictions can lead to new discoveries. When nineteenth-century astronomers observed that Uranus's location in the sky was deviating from what they had predicted, they were able to use the law of universal gravitation to infer a huge, unaccounted-for mass in the distant reaches of the solar system dragging Uranus away from its elliptical orbit. That mass turned out to be Neptune.

Taken together, the stories of Tycho, Kepler, and Newton paint a coherent picture about the relationship between prediction and understanding in science. In order to make good predictions in science, you first need to figure out how to explain the available data, facts, and theories. And if you obtain an explanation that's better than the ones that have come before, you can make a wider range of predictions about the universe than would previously have been possible.

We see this pattern, in which data enables understanding and understanding enables prediction, across

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

the sciences. Mendeleev's periodic table, which explains – albeit shallowly – similarities among the various elements, predicted the existence of never-before-seen substances. Pasteur's germ theory, which explains how people get sick, predicted that sterilization would radically reduce disease transmission in hospitals. And the theory of evolution, which arguably explains all of biology, predicted the existence of transitional species such as *Archaeopteryx* and *Tiktaalik*.

Granted, there are also countless cases where world-changing predictions were made in the absence of understanding. Perhaps the most famous is Alexander Fleming's chance discovery of the antibiotic penicillin when some of his bacterial culture plates were contaminated with a penicillin-producing mold. Observing that the mold seemed to kill nearby bacteria colonies, Fleming predicted, and successfully demonstrated, that penicillin could kill a wide variety of disease-causing bacteria. He later won the Nobel Prize in Medicine for this work, which saved innumerable lives. But no one understood how penicillin works until the 1950s and 1960s, decades after Fleming's original discovery.

Examples of prediction without understanding are common in medicine, probably because it is such a pragmatic field: Any sign of a successful disease treatment is seized upon, regardless of whether or not its mechanism is understood. Such predictions – that a selective serotonin reuptake inhibitor (SSRI) might alleviate someone's depression, for example, or that penicillin might save a child's life – are praiseworthy scientific accomplishments. But they are also fragile. Would penicillin work against some completely novel bacterium, or if it is exposed to high temperatures? Are

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

SSRIs the right choice for someone with an unusual symptom pattern? Such questions are tough to answer in the absence of a thorough understanding of how those medications work. Every time a patient with a depression spends months trying out new drugs, blindly seeking something that might work, the consequences of that lack of understanding become painfully concrete.

So if you have a collection of data about a system and you want to make some predictions – robust, varied, accurate, counterfactual predictions – you should probably gain an understanding of that system. Even if your interests are purely practical, and you're not the sort of person to ask why or how questions about the universe, you'll still be well served by finding a good explanation. Throughout most of the history of science, prediction has needed understanding.

Automating Prediction

In 2012, 244 years after Cook departed England for Tahiti, the tight link between prediction and understanding started to loosen. That year, University of Toronto graduate student Alex Krizhevsky entered the ImageNet Large Scale Visual Recognition Challenge, an annual competition for computer scientists. The goal of the competition was simple enough: Build a computer algorithm that can successfully identify the primary object in a photograph. The competition demands precise answers – "airliner," "albatross," "acorn squash" – and it's not always obvious which object in a given image is considered primary, so it's a challenging task, even for humans.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

Fei-Fei Li, a prominent Stanford computer scientist, started the ImageNet Challenge in 2010 in order to evaluate computer vision systems in a systematic way. That first year, the top algorithm achieved about 72 percent accuracy. The 2011 winner fared a bit better, reaching 74 percent accuracy, but still fell far short of human performance. Those first two victors used a tried-and-true algorithm that had been popular since the 1990s.

But Krizhevsky's 2012 entry was different. He submitted a deep neural network, a type of automated learning system that dates back to the 1960s but had never achieved widespread recognition. He called his network "AlexNet," and he trained it on two cutting-edge computer processors in his childhood bedroom. That year, the runner-up scored 74 percent, no better than the previous year's victor. But AlexNet fared far better. It scored 85 percent.

These days, Krizhevsky isn't well known outside of AI research circles, though his two AlexNet collaborators, OpenAI cofounder Ilya Sutskever and so-called Godfather of AI Geoffrey Hinton, regularly make headlines. But AlexNet's influence is indisputable. A 2025 analysis by the scientific journal *Nature* found that the AlexNet paper is the sixteenth most cited paper of all time, despite only being thirteen years old. The Computer History Museum in Mountain View, California has archived and publicly released the AlexNet source code, in recognition of the network's transformative impact on artificial intelligence.

Starting with AlexNet, almost every recent breakthrough in artificial intelligence has made use of deep neural networks. AlphaGo, the computer program that shocked

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

the world in 2016 when it defeated Go world champion Lee Sedol, is composed of two of them. Apple's FaceID and the self-driving systems in Tesla and Waymo cars all depend on them, as do image generation tools like Dall-E and Midjourney. And transformer networks, which power ChatGPT, Claude, Gemini, and every other AI chatbot product, are a specific kind of deep neural network.

Deep neural networks are just one of many approaches in machine learning, which is itself a subdiscipline of artificial intelligence. In machine learning, algorithms learn to complete tasks or identify patterns by parsing through large amounts of data, rather than depending on explicit instruction from humans. Plenty of AI systems don't make use of machine learning – Deep Blue, the IBM chess-playing computer that beat world champion Garry Kasparov in 1997, scored its possible moves according to rules that had been hard-coded by chess experts and computer scientists. But some tasks are just too tricky to boil down into a list of explicit rules; and for others, like diagnosing medical conditions based on a long list of laboratory results, scientists might not yet know what the right rules are.

So computer scientists build tools that can find patterns in data and learn to solve problems on their own, with limited human guidance. Often, the data will come in input–output pairs: images plus text labels, for example, or test results plus diagnoses. The model needs to figure out how to predict the output that goes with each input, even for examples it hasn't seen before.

Machine learning systems, then, are devices for turning data into predictions. As early as 2001, some machine

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

learning practitioners recognized how their approaches defied the conventional, understanding-dependent pipeline for making predictions in science. In a now-famous paper, the statistician Leo Breiman described the existence of "two cultures" in statistics: one culture in which statisticians try to use math to capture the *underlying relationships* that explain how input and output data fit together in nature, and a second culture wherein statisticians use any tools at their disposal to find a *reliable approach* for predicting outputs from inputs.[13] If a statistician from the first culture wanted to, say, predict the position of Mars in the sky tonight based on historical data, they might fit an equation that described Mars's orbit as an ellipse around the Sun, because they believed that Mars's orbit might really be an ellipse. A statistician from the second culture would instead throw the data at a machine learning algorithm, without spending any time contemplating the shape of Mars's orbit. With the right algorithm, they might even achieve higher predictive accuracy than their first-culture counterpart, because Kepler's laws don't accommodate the gravitational effects of the planets on each other.

Breiman advocates for the second culture, and he presents good reasons for doing so. Despite the best efforts of first-culture statisticians, he wrote, the models they build often fail to capture anything about how nature actually works. Not every statistician is a Newton or a Kepler – and, perhaps more to the point, some natural systems are still so poorly understood that even Newton could not envision a compact mathematical formula to describe how they work. If first-culture models are unlikely to reflect the structure of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

the real world, shouldn't we just choose the approach that makes the best predictions?

But as Breiman perspicaciously recognized, second-culture methods have a major shortcoming: Machine learning algorithms are often extremely difficult to understand. Worse, the most powerful algorithms are often the toughest to interpret. Breiman made this point in his paper by invoking decision trees, machine learning algorithms that learn a series of simple, binary rules (Is the patient over or under 50? Do they have diabetes?) to predict some outcome, such as the risk of heart attack. Decision trees are extremely easy to understand, but they are just okay at prediction – they can't deal with overly complex data. An alternative is to use a random forest – a big collection of decision trees that makes predictions by averaging across the predictions made by each individual tree. Random forests, Breiman explains, are powerful predictors. They are also extremely difficult to understand. It's tough to explain why a random forest made a particular prediction, when that prediction is itself the average of 100 separate predictions. Powerful machine learning algorithms, Breiman concludes, are black boxes – we may struggle to "open them up" and figure out how they work. But black boxes are often the best approach when predictive accuracy is the goal.

Breiman admitted in his paper that, as of 2001, his preference for the second culture was a minority opinion: He estimated that about 98 percent of statisticians belonged to the first culture, in which models are meant to mimic the world, and only 2 percent belonged to the machine learning culture. Since then, the landscape has changed dramatically. Deep neural networks are vastly more powerful, vastly less

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

understandable, and vastly more popular than any machine learning method was before 2012. Breiman recognized the seed of an important trend that has only recently seen its full realization.

What makes deep neural networks different from the approaches that preceded them is their ability to digest enormous volumes of data and learn how to do extraordinarily difficult tasks, with almost no human guidance. Earlier approaches needed much more hand-holding: The system that won the ImageNet competition the year before AlexNet, for example, first identified likely points of interest in the images that it processed according to hard-coded rules. Those points of interest were then fed into machine learning algorithms, which never saw the raw image data. AlexNet, however, only got the raw image data – it had to figure out how to identify points of interest on its own. And it did much better on ImageNet.

But, much as Breiman observed in the case of random forest algorithms, the very features that make deep neural networks so powerful also make them extraordinarily difficult to understand. One of those features is their scale. Deep neural networks – which, as their names suggests, do bear some coarse resemblance to the brain – are made up of enormous numbers of simulated neurons. AlexNet contains 650,000 neurons, and it's small compared to the networks popular today.

Neurons alone aren't the key determinant of a deep neural network's size. What matters more is the number of connections among those neurons. Like the cells in our brains, the neurons in a deep neural network receive information from other neurons, process that information, and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

then send the information onward to another set of neurons. All of those incoming and outgoing connections add up: AlexNet contains 600 million neuron-to-neuron connections. And those connections are how the networks learn. As AlexNet sees more and more data, some of those 600 million neuron-to-neuron connections get stronger and others get weaker. Each connection strength, or "weight," is described by a number – very positive or very negative for a strong connection, and close to zero for a weak one – which means that all of AlexNet's strategies for identifying the object in an image are somehow encoded in a list of millions of numbers. In the face of such magnitude, the interpretability challenges posed by a random forest of 100 decision trees seem quaint.

Another source of deep neural networks' power and obscurity is their depth. The neurons in a deep neural network are arranged in layers, such that neurons in the first layer receive external input (say, an image) and then send information only to the second layer, neurons in the second layer send information only to the third layer, and so on. The final layer of the network then outputs something useful – like the probability that the image contains a cat. Researchers have shown that this layer-by-layer organization allows deep networks to capture complex input–output relationships that shallower networks wouldn't be able to handle. But that depth might also make the networks harder to understand: As information flows from layer to layer, it becomes entangled in ways that are difficult to decipher.

Breiman invoked the term "black box" to describe any machine learning algorithm that is difficult to interpret, but over the past decade it has become indelibly associated with

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

deep neural networks. And science has an uneasy relationship with black boxes. In a 2005 article on scientific understanding, the philosopher Henk de Regt contrasted a prediction-granting scientific theory with an oracle whose predictions are perfect but who will not, or cannot, reveal the source of those predictions.[14] A scientific theory, such as Newtonian gravitation, can enable understanding; the oracle, on the other hand, is a black box, and black boxes can't confer understanding, no matter how accurate their predictions might be.

The deep neural networks used in science are essentially oracles – they provide excellent predictions about phenomena of scientific interest, but they can't divulge to us where those predictions came from. True, we can peek inside them to examine all of their neurons and connections in a way that we couldn't with an oracle, but that knowledge doesn't get us very far. In the face of all those weights and layers, our limited minds, which can only hold so much information at once, struggle to make any headway. If you want to know why AlexNet classified one image as a Scottish deerhound and another as a station wagon, or how it recognizes Scottish deerhounds in the first place, you're out of luck. And that ability to answer those how and why questions is key for understanding.

There are researchers who are making small advances toward attacking those questions. They use various clever techniques to try to bridge the gap between the complex inner workings of neural networks and human understanding: For example, they might try to understand the role of an individual neuron in AlexNet by generating an image that activates it as strongly as possible. Such studies have yielded interesting results. Neurons in the network's early layers tend to respond

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

to images full of geometric patterns, whereas neurons in later layers respond best to images that contain identifiable objects, like eyes and buildings. That research does give us some purchase on how AlexNet-like networks work – they seem to analyze the basic shapes in an image before building them up into more recognizable objects. But that's still not enough information to explain how and why AlexNet classifies any particular image in the way that it does.

Even if we get a better sense of how and why deep neural networks function in the ways that they do, we might not improve our understanding of the universe very much at all. To see this, imagine an oracle who is excellent at predicting storms. If you learned that they were making their predictions using a barometer, you would understand their process much better – you'd be able to answer how and why questions about the predictions that they made. But your understanding of storms might not advance at all, at least not without breaking down the barometer to figure out how *it* works. Much like an oracle with a barometer, deep neural networks may make predictions about phenomena using methods that don't have anything obvious to do with how or why those phenomena occur. Deep neural networks are excellent at recognizing intricate patterns in datasets that a human would never detect. That makes them stellar predictors, but it might also render them poor explainers.

The Future of Understanding

Deep neural networks have made data far more powerful, but, contra Anderson, they have not ended science. Were

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

some future AI-enhanced human to page through all of the scientific literature produced in the last fifteen years, they would likely conclude that deep neural networks have in fact acted as a potent accelerant for science. The clearest example is AlphaFold, a deep neural network that predicts the three-dimensional structure of proteins and is the focus of the next chapter. AlphaFold predictions are often accurate enough to substitute for experimental results, and they can be obtained far more quickly, so any molecular biologist who wants to understand how a particular protein works can compound their effectiveness by using AlphaFold predictions.

But AlphaFold isn't omnipotent: We still need human scientists to look at AlphaFold-predicted structures and infer what they might mean for how proteins function within the cell. In the grand scheme of science, the problems that specialized deep neural networks can fully solve are few and far between. They can only be used in settings where large volumes of data are available, so some disciplines remain effectively untouched; and even in fields where neural networks have automated some predictions, such as molecular biology, many other problems demand human understanding. But in areas where deep learning is most influential, we are seeing something new and important: Understanding is no longer automatically justified by predictive goals, and the intellectual view needs to be advocated for on its own merits.

For scientists, that may not be too tall an order. In 2025 I spoke with Henk de Regt, who now directs the Institute for Science in Society at Radboud University in the Netherlands. Our hour-long conversation centered on AI

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

and scientific understanding, a topic that has lately become a major focus of his research. Given that de Regt has spent practically his entire philosophical career working on scientific understanding, I had expected that he might meet the growing prominence of deep neural networks in science with some trepidation. But he doesn't envision an end to understanding on the horizon just yet. "As long as human scientists remain part of the game," he says, "I think they will want understanding."

Based on the many conversations I have had with scientists while reporting this book, I think de Regt is correct. It's certainly true that some scientists choose their careers with the primary goal of curing diseases or developing new technologies rather than furthering human understanding – but even if a particular scientist prioritizes the practical view over the intellectual view, they probably subscribe to both in some degree. Both views are deeply embedded in scientific practice, and someone who doesn't care much about advancing their understanding of the world probably wouldn't get far as a researcher.

That said, the scientific community doesn't operate in a vacuum. These days, those who fund science, whether governments or private foundations, are likely to espouse the practical view: They want to see results. And funding pressure are getting more intense, at least in the United States: In 2025, the Trump administration dramatically reduced grant-making by both the National Institutes of Health and the National Science Foundation. Simultaneously, the administration has enthusiastically supported the use of AI in science through initiatives such as the Genesis Mission,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

which aims to create new scientific AI systems trained on federal datasets.

While public funding for research is in disarray, private companies are rapidly developing new, blockbuster AI tools, such as ChatGPT and Claude, that are just now starting to be integrated into the scientific process. Whereas a purpose-built scientific neural network such as AlphaFold can perform a single task extremely well, these general-purpose AI systems promise to take on a wide range of scientific activities. If such systems do prove themselves capable of formulating hypotheses, devising experiments, and assimilating new discoveries, they could potentially render human understanding superfluous in a much greater variety of scientific contexts – at least according to the practical view.

Science thus stands poised at a precipice, and it's an apt time to take stock. In the next three chapters, I will document how researchers in three fields – molecular biology, particle physics, and neuroscience – are navigating a scientific landscape occupied by oracles. Often, these oracles act as effective assistants, enabling quicker progress on some problems even as they marginalize others. Some researchers are working to uncover the oracles' secrets, in the hope that they might enable new insights. In some fields, certain problems are being quietly deemphasized or prematurely declared solved; in others, the ideal of what a solution might look like is shifting. In small but meaningful ways, the practical view and the intellectual view of science are starting to peel apart.

I have decided to focus these chapters – the bulk of this book – on deep neural networks that have been designed

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E

to address specific scientific questions rather than the large language models (LLMs) that power popular AI chatbots such as ChatGPT and Claude. That might seem like a strange choice – for many people, "AI" and "LLM" are synonyms. But I think it's the most sensible way to attack the question of what AI means for science, at least for the time being. While there is plenty of discussion of what LLMs might have to offer science, and plenty of researchers and entrepreneurs who are working to make those visions a reality, those projects aren't yet mature, and the situation on the ground is changing rapidly. Specialized deep neural networks, on the other hand, have been part of science for over a decade, and many scientists have fully integrated them into their work. By focusing on these widespread and now-familiar tools, we can uncover legible signs of things to come.

In a sense, then, this book functions as an advance alert, or a call to attention. The seeds that could one day enable science to transition into a wholly practical regime, in which human understanding is construed as an unnecessary indulgence, have been sown. This world is not necessarily coming soon – indeed, it may never arrive – but it behooves us to consider the possible consequences now. Perhaps a full-throatedly practical science is desirable: If we can develop cures for diseases and clean energy sources without waiting for limited human minds to understand the nature of those solutions, so much the better. Or perhaps surrendering the possibility of grasping the universe's workings would be too dear a loss. I won't be answering that question here. My goal is simply to get the conversation started.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 08:58:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/AF4216986A16E5B3A770D149A983771E