HOW AI WORKS

FROM SORCERY TO SCIENCE

RONALD T. KNEUSEL

PRAISE FOR
HOW AI WORKS

"A must-read for anyone wishing to dig into AI without getting lost in the weeds. Kneusel has succeeded in explaining how AI works to a layperson like myself."

—KENNETH GASS, HONORARY CURATOR OF GEOLOGY, MILWAUKEE PUBLIC MUSEUM

"*How AI Works* is a friendly and personal peek behind the curtain of modern AI. Ronald T. Kneusel tells the story of how the field grew, and surveys the ideas that are powering the AI revolution. From this book, you'll learn not only how AI works today, but its limits, its capabilities, and where it might take us tomorrow."

—ANDREW GLASSNER, AUTHOR OF *DEEP LEARNING: A VISUAL APPROACH*

"*How AI Works* is a tour de force of the rich history of artificial intelligence, from the early perceptions and symbolic systems to large language models such as ChatGPT. For beginners, it demystifies AI and is a perfect resource to get up to date with more than six decades of research and development. For those versed in AI, it serves as an invaluable tool to fill knowledge gaps. Even AI experts will gain a fresh perspective, enhancing their understanding and ability to articulate complex concepts."

—BEN DICKSON, SOFTWARE ENGINEER, EDITOR OF TECHTALKS

"After reading this book I have a better understanding of the ML tools I have already used in my work, and a new appreciation and insight to how Large Language Models, and future AI, will likely change the domains in which I work. I recommend this book to anyone who works with software systems, including management, and anyone who just wants to know what AI actually does under the hood."

—DANIEL KOSEY, CISSP, CYBERSECURITY ENGINEER

HOW AI WORKS

From Sorcery to Science

by Ronald T. Kneusel

no starch
press®

San Francisco

To Frank Rosenblatt—he saw it coming.

About the Author

Ronald T. Kneusel has been working with machine learning in industry since 2003 and completed a PhD in machine learning at the University of Colorado, Boulder, in 2016. Ron has written five other books: *Practical Deep Learning: A Python-Based Introduction* (No Starch Press, 2021), *Math for Deep Learning: What You Need to Know to Understand Neural Networks* (No Starch Press, 2021), *Strange Code: Esoteric Languages That Make Programming Fun Again* (No Starch Press, 2022), *Numbers and Computers* (Springer, 2017), and *Random Numbers and Computers* (Springer, 2018).

About the Technical Reviewer

Alex Kachurin is a data science and machine learning professional with more than 15 years of experience in the field. He earned an MS in computer vision from the University of Central Florida in 2010.

CONTENTS

ACKNOWLEDGMENTS

PREFACE

Many books teach you how to do artificial intelligence (AI). Similarly, many popular books tell you about AI. However, what seems to be missing is a book that teaches you how AI works at a conceptual level. AI isn't magic; you can understand what it's doing without burying yourself in complex mathematics.

This book fills that void with a math-free explanation of how AI works. While some books are down in the weeds and others offer a bird's-eye view, this book is at treetop level. It aims to provide you with enough detail to understand the approach without getting bogged down in nitty-gritty mathematics. If that piques your interest, I invite you to read on.

You'll run across places where **** appears throughout the book. These markers highlight a shift in the topic or a transition point. In a textbook, **** would indicate a new section, but this isn't a textbook, nor do I want it to feel like one; so, instead of sections and subsections, I'll use asterisks to warn you that a change is coming. Like this …

****

I first learned about artificial intelligence in 1987, in an undergraduate course of the same name. What people typically mean by *AI* has changed somewhat over the

intervening decades. Still, the goal remains the same: to mimic intelligent behavior in a machine.

Few people in the 1980s had any reason to learn about AI, if they were even aware of it. AI had minimal impact on their daily lives, beyond the occasional renegade computer in science fiction TV shows and movies like *Star Trek* or *WarGames*, to say nothing of the relentless and terrifying *Terminator*.

However, the 1980s are long gone, current retro fashion trends notwithstanding, and AI is everywhere. It affects our lives in numerous ways every day, from phones telling us to drive here and not there, to labeling friends and family in pictures, to the articles and ads fed to us continuously online, like it or not. And this is to say nothing of the recent AI explosion involving large language models, which many interpret as "true AI" at last.

AI is also there behind the scenes in ways we seldom realize: airline flight planning, shipping and logistics, factory automation, satellite imaging of the earth, and helping your doctor decide if that lump is cancer, to name a few.

Why learn about AI now?

This book answers that question by explaining what happened, when it happened, why it happened, and, most importantly, how it happened—all without hype or a single mathematical equation. Frankly, the reality behind the AI revolution is impressive enough; the hype is unnecessary.

At this point, I feel some words about me are in order. After all, I'm asking you to join me on a journey through the world of AI, so it's reasonable to wonder about your guide. I certainly would.

As mentioned earlier, I was introduced to AI in the late 1980s. I began working in AI, in the subfield known as *machine learning*, in 2003, applying machine learning models to intravascular ultrasound images.

I first heard of deep learning in 2010. *Deep learning* is a subfield of machine learning. I'll clarify the difference between deep learning, machine learning, and artificial intelligence in Chapter 1, but for now you can think of them as the same thing.

In 2012, AI burst onto the scene—or at least into the news—with the advent of what came to be called AlexNet and a curious experiment at Google involving computers that learned to identify cats in YouTube videos. I was in the room at the 2012 International Conference on Machine Learning in Edinburgh, Scotland, when Google presented its paper. It was standing room only for the conference's 800 or so attendees.

In 2016, I completed a PhD in computer science specializing in AI at the University of Colorado, Boulder, under the direction of Michael Mozer. I've worked in AI daily since then, primarily in the defense industry, with a short break in 2016 to help co-found a medical AI startup.

After AlexNet, things changed quickly, as seemingly monthly some new AI-related "miracle" appeared in the academic literature, if not on the evening news. The only way to keep up was to attend conferences multiple times per year; waiting for results to appear in an academic journal was pointless, as the field was progressing too rapidly for the typically slow pace of academic publishing.

I'm writing this preface in November 2022 at the NeurIPS conference. NeurIPS is arguably the premier AI conference (no hate emails, please!), and this is the first time it's been held in person since the COVID-19 pandemic. Attendance is high, though perhaps not as high as at the 2019 conference, for which a lottery was held to determine which 13,500 people could attend. The fact that conference attendance has blossomed from a few hundred to over 10,000 in a decade tells us how important AI research has become.

The names of the tech industry leaders who support these conferences, which are prime hunting grounds for graduate students, also reveal the significance of AI. You'll find expo booths for Google, DeepMind (also Google), Meta (read: Facebook), Amazon, Apple, and others. AI drives much of what these companies do. AI is big bucks. AI runs on data, and these companies gobble up all the data we freely give them in exchange for their services.

By the end of the book, you'll understand what AI is doing under the hood (or bonnet, if you prefer). Ultimately, it isn't all that difficult to comprehend, though the devil is definitely in the details.

The book proceeds as follows:

Chapter 1, And Away We Go: An AI Overview We dive in with a quick overview of AI essentials and a basic example.

Chapter 2, Why Now? A History of AI AI didn't just fall from the sky. This chapter gives you AI's backstory and clarifies why the revolution is happening now.

Chapter 3, Classical Models: Old-School Machine Learning Modern AI is all neural networks, but to understand what neural networks are doing, it helps to understand the models that came before.

Chapter 4, Neural Networks: Brain-Like AI If you want to know what a neural network is, how it's trained, and how it's used, then this chapter is for you.

Chapter 5, Convolutional Neural Networks: AI Learns to See Much of the power of modern AI comes from learning new ways to represent data. If that sentence has no meaning for you, this chapter will help.

Chapter 6, Generative AI: AI Gets Creative Traditional supervised machine learning models attach labels to inputs. Generative AI produces novel output, including text, images, and even video. This chapter explores two popular approaches: generative adversarial networks (GANs) and diffusion models. GANs provide the intuition we need to explore diffusion models and, in Chapter 7, large language models (LLMs). Diffusion models are adept at producing detailed, photorealistic images and videos from text prompts.

Chapter 7, Large Language Models: True AI at Last? OpenAI's fall 2022 release of its large language model, ChatGPT, might very well have ushered in the era of true AI. This chapter explores LLMs: what they are, how they work, and the claim that they are something new and disruptive.

Chapter 8, Musings: The Implications of AI The advent of large language models has altered the AI landscape. This chapter muses on the implications.

At the end of the book, you'll find a collection of additional resources to explore, should the AI bug bite and you want to learn more. Personally, and admittedly with bias, I recommend my books *Practical Deep Learning: A Python-Based Introduction* (2021) and *Math for Deep Learning: What You Need to Know to Understand Neural Networks* (2021), both available from No Starch Press. They will give you what you need to go from reading about how AI works conceptually to "doing" AI.

Finally, as you read, you'll notice that specific phrases in the text are *emphasized*. Definitions for many of these emphasized words and phrases are found in the glossary at the end of the book. Like every field, AI has its jargon. Keeping all the terms in your head is burdensome, hence the glossary to help you remember them.

I'm a real person. I know because I can successfully identify and click images of trains and traffic lights. If you have comments or questions about the material in this book, I want to hear from you. Please email me at *rkneuselbooks@gmail.com*.

Now, if you're ready, away we go.

1
AND AWAY WE GO: AN AI OVERVIEW

Artificial intelligence attempts to coax a machine, typically a computer, to behave in ways humans judge to be intelligent. The phrase was coined in the 1950s by prominent computer scientist John McCarthy (1927–2011).

This chapter aims to clarify what AI is and its relationship to *machine learning* and *deep learning*, two terms you may have heard in recent years. We'll dive in with an example of machine learning in action. Think of this chapter as an overview of AI as a whole. Later chapters will build on and review the concepts introduced here.

****

Computers are *programmed* to carry out a particular task by giving them a sequence of instructions, a *program*, which embodies an *algorithm*, or the recipe that the program causes the computer to execute.

The word *algorithm* is cast about often these days, though it isn't new; it's a corruption of *Al-Khwarizmi*, referring to ninth-century Persian mathematician Muhammad ibn Musa al-Khwarizmi, whose primary gift to the world was the mathematics we call *algebra*.

****

Let's begin with a story.

Tonya owns a successful hot sauce factory. The hot sauce recipe is Tonya's own, and she guards it carefully. It's literally her secret sauce, and only she understands the process of making it.

Tonya employs one worker for each step of the hot sauce–making process. These are human workers, but Tonya treats them as if they were machines because she's worried they'll steal her hot sauce recipe—and because Tonya is a bit of a monster. In truth, the workers don't mind much because she pays them well, and they laugh at her behind her back.

Tonya's recipe is an algorithm; it's the set of steps that must be followed to create the hot sauce. The collection of instructions Tonya uses to tell her workers how to make the hot sauce is a program. The program embodies the algorithm in a way that the workers (the machine) can follow step by step. Tonya has programmed her workers to implement her algorithm to create hot sauce. The sequence looks something like this:

Tonya's recipe ⟶ Tonya's instructions ⟶ The workers
(the algorithm)   (the program)   (the machine)
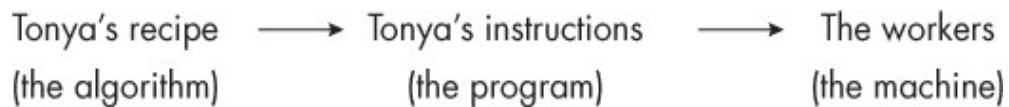
There are a few things to note about this scenario. First, Tonya is definitely a monster for treating human beings as machines. Second, at no point in the process of making hot sauce does any worker need to understand why they do what they do. Third, the programmer (Tonya) knows why the machine (the workers) does what it does, even if the machine doesn't.

****

What I've just described is how we've controlled virtually all computers, going back to the first conceptual machines envisioned by Alan Turing in the 1930s and even earlier to the 19th-century Analytical Engine of Charles Babbage. A human conceives an algorithm, then translates

that algorithm into a sequence of steps (a program). The machine executes the program, thereby implementing the algorithm. The machine doesn't understand what it's doing; it's simply performing a series of primitive instructions.

The genius of Babbage and Turing lay in the realization that there could be a general-purpose machine capable of executing arbitrary algorithms via programs. However, I would argue that it was Ada Lovelace, a friend of Babbage's often regarded as the world's first programmer, who initially understood the far-reaching possibilities of what we now call a computer. We'll talk more about Turing, Babbage, and Lovelace in Chapter 2.

NOTE *In Lovelace's day, a "computer" was not a machine but a human being who calculated by hand. Hence, Babbage's Engine was a mechanical computer.*

Let's take a moment to explore the relationship between the terms *AI*, *machine learning*, and *deep learning*. On the one hand, all three have become synonymous as referring to modern AI. This is wrong, but convenient. Figure 1-1 shows the proper relationship between the terms.
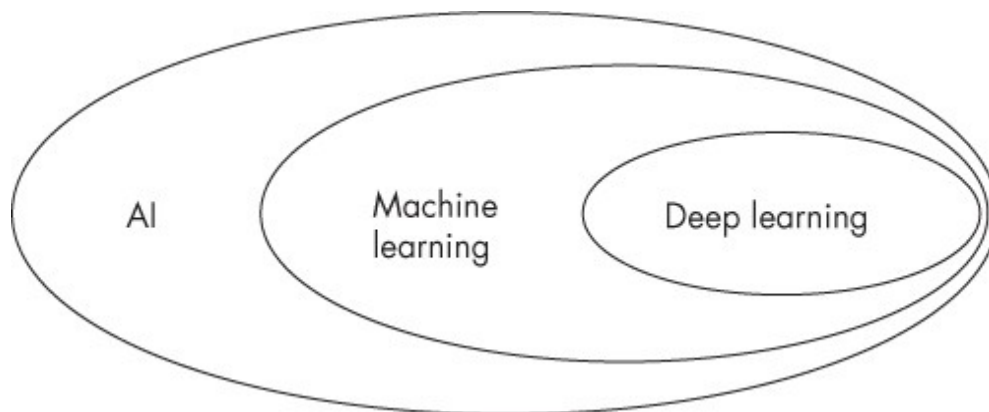
Figure 1-1: The relationship between artificial intelligence, machine learning, and deep learning

Deep learning is a subfield of machine learning, which is a subfield of artificial intelligence. This relationship implies that AI involves concepts that are neither machine learning nor deep learning. We'll call those concepts *old-school AI*,

which includes the algorithms and approaches developed from the 1950s onward. Old-school AI is not what people currently mean when discussing AI. Going forward, we'll entirely (and unfairly) ignore this portion of the AI universe.

Machine learning builds models from data. For us, a *model* is an abstract notion of something that accepts inputs and generates outputs, where the inputs and outputs are related in some meaningful way. The primary goal of machine learning is to condition a model using *known* data so that the model produces meaningful output when given *unknown* data. That's about as clear as muddy water, but bear with me; the mud will settle in time.

Deep learning uses large models of the kind previously too big to make useful. More muddy water, but I'm going to argue that there's no strict definition of deep learning other than that it involves neural networks with many layers. Chapter 4 will clarify.

In this book, we'll be sloppy but in accord with popular usage, even by experts, and take "deep learning" to mean large neural networks (yet to be formally defined), "machine learning" to mean models conditioned by data, and "AI" to be a catchall for both machine learning and deep learning—remembering that there is more to AI than what we discuss here.

Data is everything in AI. I can't emphasize this enough. Models are blank slates that data must condition to make them suitable for a task. If the data is bad, the model is bad. Throughout the book, we'll return to this notion of "good" and "bad" data.

For now, let's focus on what a model is, how it's made useful by conditioning, and how it's used after conditioning. All this talk of conditioning and using sounds dark and sinister, if not altogether evil, but, I assure you, it's not, even though we have ways of making the model talk.

****

A machine learning model is a black box that accepts an input, usually a collection of numbers, and produces an output, typically a label like "dog" or "cat," or a continuous value like the probability of being a "dog" or the value of a house with the characteristics given to the model (size, number of bathrooms, ZIP code, and so on).

The model has *parameters*, which control the model's output. Conditioning a model, known as *training*, seeks to set the model's parameters in such a way that they produce the correct output for a given input.

Training implies that we have a collection of inputs, and the outputs the model should produce when given those inputs. At first blush, this seems a bit silly; why do we want the model to give us an output we already have? The answer is that we will, at some future point, have inputs for which we don't already have the output. This is the entire point of making the model: to use it with unknown inputs and to believe the model when it gives us an output.

Training uses the collection of known inputs and outputs to adjust the model's parameters to minimize mistakes. If we can do that, we begin to believe the model's outputs when given new, unknown inputs.

Training a model is fundamentally different from programming. In programming, we implement the algorithm we want by instructing the computer step by step. In training, we use data to teach the model to adjust its parameters to produce correct output. There is no programming because, most of the time, we have no idea what the algorithm should be. We only know or believe a relationship exists between the inputs and the desired outputs. We hope a model can approximate that relationship well enough to be useful.

It's worth remembering the sage words of British statistician George Box, who said that all models are wrong, but some are useful. At the time, he was referring to