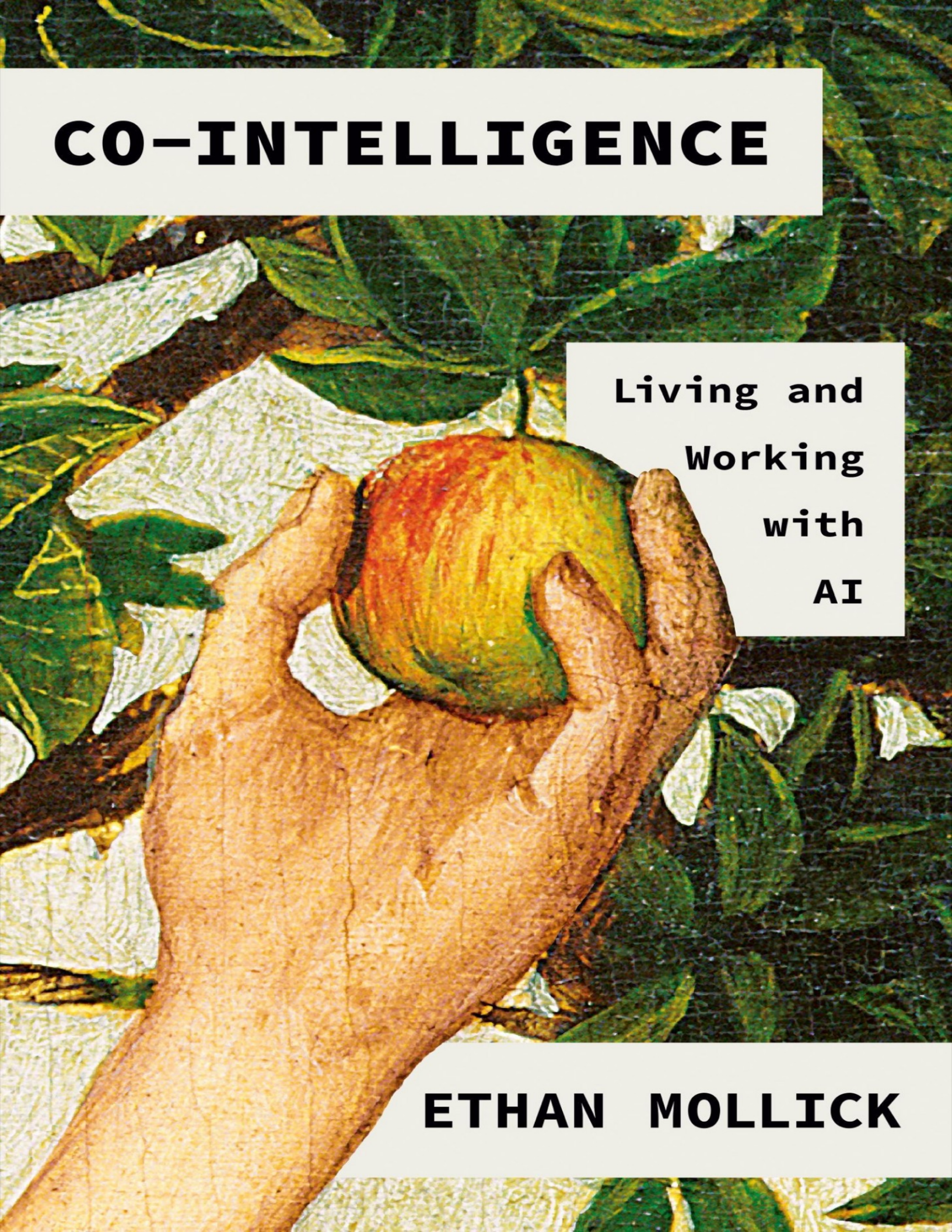
CO-INTELLIGENCE
Living and Working with AI
ETHAN MOLLICK

# CO-INTELLIGENCE

## Living and Working with AI

ETHAN
MOLLICK

Portfolio | Penguin

Portfolio / Penguin

An imprint of Penguin Random House LLC

penguinrandomhouse.com



LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Names: Mollick, Ethan, 1975– author.

Title: Co-intelligence: living and working with AI / Ethan Mollick.

Other titles: Cointelligence

Description: [New York]: Portfolio/Penguin, [2024] | Includes bibliographical references.

Identifiers: LCCN 2023049476 (print) | LCCN 2023049477 (ebook) | ISBN 9780593716717 (hardcover) | ISBN 9780593852507 (international edition) | ISBN 9780593716724 (ebook)

Subjects: LCSH: Expert systems (Computer science)—Social aspects. | Artificial intelligence—Social aspects. | Artificial intelligence—Educational applications. | Labor—Effect of technological innovations on. | Education—Effect of technological innovations on.

Classification: LCC QA76.76.E95 M655 2024 (print) | LCC QA76.76.E95 (ebook) | DDC 303.48/34—dc23/eng/20240209

LC record available at https://lccn.loc.gov/2023049476

LC ebook record available at https://lccn.loc.gov/2023049477

Cover design: Brian Lemus

Cover art: Detail of *The Fall, 1479* by Hugo van der Goes (oil on panel) / Photo © Gordon Roberton Photography Archive/Bridgeman Images

Book design by Chris Welch

All AI-generated images and text are clearly noted.

While the author has made every effort to provide accurate internet addresses at the time of publication, neither the publisher nor the author assumes any responsibility for errors or for changes that occur after publication. Further, the publisher does not have any control over and does not assume any responsibility for author or third-party websites or their content.

pid_prh_6.3_146644690_co_ro

*To Lilach Mollick*

# Contents

# Introduction

# THREE SLEEPLESS NIGHTS

I believe the cost of getting to know AI—really **getting to know** AI—is at least three sleepless nights.

After a few hours of using generative AI systems, there will come a moment when you realize that Large Language Models (LLMs), the new form of AI that powers services like ChatGPT, don't act like you expect a computer to act. Instead, they act more like a person. It dawns on you that you are interacting with something new, something alien, and that things are about to change. You stay up, equal parts excited and nervous, wondering: What will my job be like? What job will my kids be able to do? Is this thing thinking? You go back to your computer in the middle of the night and make seemingly impossible requests, only to see the AI fulfill them. You realize the world has changed in fundamental ways and that nobody can really tell you what the future will look like.

Though I am not a computer scientist, I am an academic studying innovation who has long been involved in work on the applications of AI, especially for learning. Over the years, AI has promised much more than it has delivered. For decades, AI research has always seemed to be on the edge of a massive breakthrough, but most practical uses, from self-driving cars to personalized tutoring, always advanced grindingly slowly. During this time, I kept experimenting with AI tools, including OpenAI's GPT models, figuring out ways to incorporate them into my work, and assigning my students to

use AI in class. So my sleepless nights came early, just after the release of ChatGPT in November 2022.

After only a couple of hours, it was clear that something huge had shifted between previous iterations of GPT and this new one. Four days after the AI was launched, I decided to demonstrate this new tool to my undergraduate entrepreneurship class. Barely anyone had heard of it. In front of my students, I put on a show, demonstrating how AI can help generate ideas, write business plans, turn those business plans into poems (not that there is a lot of demand for that), and generally fill the role of company cofounder. By the end of the class, one of my students, Kirill Naumov, had created a working demo for his entrepreneurship project—a Harry Potter–inspired moving picture frame that reacted to people walking near it—using a code library he had never used before, in less than half the time it would otherwise have taken. He had venture capital scouts reaching out to him by the end of the next day.

Within two days of introducing students to AI, several told me they used ChatGPT to explain confusing concepts to them "like they were ten years old." They stopped raising their hands as much—why expose themselves in class when they could just ask the AI later? And every essay was suddenly written with perfect grammar (though references were often wrong and the final paragraph tended to start with "In conclusion"—a telltale sign of early ChatGPT writing, since fixed). But the students weren't just excited, they were nervous. They wanted to know the future.

Some of them asked me what it means for their preferred careers ("Should I become a radiologist if AI can do a lot of the work?" "Will writing marketing copy still be a good job five years from now?"). Others asked when this technological development would stop, or if it would stop. Someone even asked if Artificial General Intelligence (AGI), the hypothetical machine that would be smarter than a human, might arrive before graduation.

I didn't have any answers then (though I have a few more now), yet I recognized that I wasn't immune to these changes. So I decided to get hands-on and put the AI through its paces. Much of my work over the past five

years has been building simulations that teach business skills. It has taken tremendous investment from my university, and the efforts of a dozen talented people, to build these elaborate digital experiences. After thousands of hours, the games were great: using thousands of lines of code, we could do elaborate learning simulations that helped teach skills like negotiation. But I decided to type something into ChatGPT:

---

> You will be my negotiation teacher. You will simulate a detailed scenario in which I have to engage in a negotiation. You will fill the role of one party, I will fill the role of the other. You will ask for my response in each step of the scenario and wait until you receive it. After getting my response, you will give me details of what the other party does and says. You will grade my response and give me detailed feedback about what to do better using the science of negotiation. You will give me a harder scenario if I do well, and an easier one if I fail.

---

ChatGPT wrote back:

**Sure, I'd be happy to help you practice negotiations through a simulation exercise! Let's start with a simple scenario. You are a salesperson trying to sell 100 pens to a customer. The pens are usually $1, but the customer is trying to negotiate the price down. The customer starts by offering to buy the pens for $0.50 each. How do you respond?**

---

And it proceeded to take me through an entire simulation that, while not perfect, did 80 percent of what took our team months to do. Based on just a paragraph-long prompt from me!

My sleepless nights began.

And to be honest, they haven't ended. Every week, it seems like AI serves up a new miracle—or a new concerning development. I soon found myself arguing with a chatbot that accused me of being unethical when I asked it to help me write computer code. I discovered that I could create beautiful images by giving the right words to a machine. I could code in Python, a language I never learned. I discovered that a machine could do large amounts of my job . . . and maybe lots of other people's jobs. I discovered something remarkably close to an alien co-intelligence, one that can interact well with humans, without being human or, indeed, sentient. I think we will all have our three sleepless nights soon.

And sleep-deprived as I now am, I keep going back to my students' questions: What does this technology mean for the future of work and education? Things are happening so quickly that it is hard to be sure, but we can begin to see its outlines.

AI is what those of us who study technology call a General Purpose Technology (ironically, also abbreviated GPT). These advances are once-in-

a-generation technologies, like steam power or the internet, that touch every industry and every aspect of life. And, in some ways, generative AI might even be bigger.

General Purpose Technologies typically have slow adoption, as they require many other technologies to work well. The internet is a great example. While it was born as ARPANET in the late 1960s, it took nearly three decades to achieve general use in the 1990s, with the invention of the web browser, the development of affordable computers, and the growing infrastructure to support high-speed internet. It was fifty years before smartphones enabled the rise of social media. And many companies have not even fully embraced the internet: making a business "digital" is still a hot topic of discussion at business school, especially as many banks still use mainframe computers. And previous General Purpose Technologies have similarly taken many decades from development until they were useful. Consider computers, another transformative technology. Early computers improved quickly, thanks to Moore's Law, the long-standing trend that the capability of computers doubles every two years. But it still took decades for computers to start appearing at businesses and schools because, even with their fast rate of increasing ability, they were starting from a very primitive beginning. Yet Large Language Models proved incredibly capable within a few years of their invention. They've also been adopted by consumers very quickly; ChatGPT reached 100 million users faster than any previous product in history, driven by the fact that it was free to access, available to individuals, and incredibly useful.

They are also getting better. The size of these models is increasing by an order of magnitude a year, or even more, so their capability is also improving. Even though that progress will likely slow, it is happening at a pace that dwarfs any other major technology, and LLMs are just one of a set of potential machine learning technologies powering the new wave of AI. Even if AI development were to stop as I was finishing this sentence, it would still transform our lives.

Finally, as great as previous General Purpose Technologies were, their impact on work and education may actually be less than the impact of AI. Where previous technological revolutions often targeted more mechanical and repetitive work, AI works, in many ways, as a co-intelligence. It augments, or potentially replaces, human thinking to dramatic results. Early studies of the effects of AI have found it can often lead to a 20 to 80 percent improvement in productivity across a wide variety of job types, from coding to marketing. By contrast, when steam power, that most fundamental of General Purpose Technologies, the one that created the Industrial Revolution, was put into a factory, it improved productivity by 18 to 22 percent. And despite decades of looking, economists have had difficulty showing a real long-term productivity impact of computers and the internet over the past twenty years.

Plus, General Purpose Technologies aren't just about work; they touch every aspect of our lives. They change how we teach, entertain ourselves, interact with other people, and even our sense of self. Schools are in an uproar over the future of writing, based on the first generation of AIs, and AI tutors may finally radically change how we educate students. AI-driven entertainment allows for stories to be personalized to us and is sending shock waves through Hollywood. And AI-driven misinformation is already flowing through social networks in ways that are difficult to detect and deal with. Things are about to get very strange; in fact, if you know where to look, they are already getting strange.

And all of this ignores the larger issue, the alien in the room. We have created something that has convinced many smart people that it is, in some way, the spark of a new form of intelligence. An AI that has blown through both the Turing Test (Can a computer fool a human into thinking it is human?) and the Lovelace Test (Can a computer fool a human on creative tasks?) within a month of its invention, an AI that aces our hardest exams, from the bar exam to the neurosurgery qualifying test. An AI that maxes out our best measures for human creativity and our best tests for sentience.

Even weirder, it is not entirely clear why the AI can do all these things, even though we built the system and understand how it technically works.

No one really knows where this is all heading, including me. Yet, despite not having definitive answers, I think I can be a useful guide. I have found myself to be an influential voice on the implications of AI, particularly through my newsletter, *One Useful Thing*, even though I am not a computer scientist myself. Indeed, I think that one of my advantages in understanding AI is that, as a professor at Wharton, I have long studied and written about how technologies are *used*. As a result, my coauthors and I have published some of the first research on AI in education and in business, and we have been experimenting with practical uses of AI in ways that major AI companies have cited as examples. I regularly speak with organizations, companies, and government agencies, as well as with many AI experts, to understand the world we are making. I also attempt to keep up with the flood of research in the field, much of it in the form of scientific working papers that have not yet gone through the long process of peer review but still offer valuable data about this new phenomenon (I will be citing a lot of this early work in the book to help fill in the picture of where we are headed, but it is important to realize that the field is evolving rapidly). Based on all these conversations and papers, I can assure you that there is nobody who has the complete picture of what AI means, and even the people making and using these systems do not understand their full implications.

So I want to try to take you on a tour of AI as a new thing in the world, a co-intelligence, with all the ambiguity that the term implies. We have invented technologies, from axes to helicopters, that boost our physical capabilities; and others, like spreadsheets, that automate complex tasks; but we have never built a generally applicable technology that can boost our intelligence. Now humans have access to a tool that can emulate how we think and write, acting as a co-intelligence to improve (or replace) our work. But many of the companies developing AI are going further, hoping to create a sentient machine, a truly new form of co-intelligence that would coexist

with us on Earth. To get a handle on what this means, we need to start from the beginning, with a very basic question: What is AI?

So we are going to start there, discussing the technology of Large Language Models. That will give us a basis for thinking about how we, as humans, can best work with these systems. After that, we can dive into how AI can change our lives by acting as a coworker, a teacher, an expert, and even a companion. Finally, we can turn to what this might mean for us, and what it means to think together with an alien mind.

# PART I

# 1 CREATING ALIEN MINDS

Talking about AI can be confusing, in part because AI has meant so many different things and they all tend to get muddled together. Siri telling you a joke on command. The Terminator crushing a skull. Algorithms predicting credit scores.

We've long had a fascination with machines that can think. In 1770, the invention of the first mechanical chess computer stunned those who saw it—a chessboard set upon an elaborate cabinet, with its chess pieces manipulated by a robot dressed as an Ottoman wizard. It toured the world from 1770 to 1838. The machine, also known as the Mechanical Turk, beat Ben Franklin and Napoleon in chess matches and led Edgar Allan Poe to speculate on the possibility of artificial intelligence upon seeing it in the 1830s. It was all a lie, of course—the machine cleverly hid a real chess master inside its fake gears, but our ability to believe that machines might be able to think fooled many of the best minds in the world for three quarters of a century.

Fast-forward to 1950, when a toy and a thought experiment, each developed by a different genius of the still-developing field of computer science, led to a new conception of artificial intelligence. The toy was a jury-rigged mechanical mouse called Theseus, developed by Claude Shannon, an inventor, prankster, and the greatest information theorist of the twentieth century. In a 1950 film, he revealed that Theseus, powered by repurposed telephone switches, could navigate through a complex maze—the first real example of machine learning. The thought experiment was the imitation

game, where computer pioneer Alan Turing first laid out the theories about how a machine could develop a level of functionality sufficient to mimic a person. While computers were a very new invention, Turing's influential paper helped kick off the nascent field of artificial intelligence.

Theories alone were not enough, and a handful of early computer scientists started working on programs that pushed the boundaries of what was soon called artificial intelligence, a term invented in 1956 by John McCarthy of MIT. Progress was initially rapid as computers were programmed to solve logic problems and play checkers—leading researchers expected an AI to beat grandmasters in chess within a decade. But hype cycles have always plagued AI, and as these promises went unfulfilled, disillusionment set in, one of many "AI winters" in which AI progress stalls and funding dries up. Other boom-and-bust cycles followed, each boom accompanied by major technological advances, such as artificial neural networks that mimicked the human brain, followed by collapse as AI could not deliver on expected goals.

The latest AI boom started in the 2010s with the promise of using machine learning techniques for data analysis and prediction. Many of these applications used a technique called supervised learning, which means these forms of AI needed labeled data to learn from. Labeled data is data that has been annotated with the correct answers or outputs for a given task. For example, if you want to train an AI system to recognize faces, you need to provide it with images of faces that have been labeled with the names or identities of the people in them. This phase of AI was the domain of larger organizations that had vast amounts of data. They used these tools as powerful prediction systems, whether optimizing shipping logistics or guessing what kind of content to show you based on your browsing history. You might have heard the buzzwords *big data* or *algorithmic decision-making* describing these kinds of uses. Consumers mostly saw the benefits of machine learning when these techniques were integrated into tools such as voice recognition systems or translation apps. *AI* was a poor (albeit marketing-friendly) label for what this sort of software did, since there was

very little about these systems that actually seemed intelligent or clever, at least in the ways humans are intelligent and clever.

To see one example of how this sort of AI works, picture a hotel attempting to forecast its demand for the upcoming year, armed with nothing but existing data and a simple Excel spreadsheet. Before predictive AI, hotel owners would often be left playing a guessing game, trying to predict demand while grappling with inefficiencies and wasted resources. With this form of AI, they could instead input a wealth of data—weather patterns, local events, and competitor pricing—and generate far more accurate predictions. The results were a more efficient operation and, ultimately, a more profitable business. Before machine learning and natural language processing became mainstream, organizations focused on being correct on average—a rather rudimentary approach by today's standards. With the introduction of AI algorithms, the focus shifted to statistical analysis and minimizing variance. Instead of being right on average, they could be right for each specific instance, leading to more accurate predictions that revolutionized many back-office functions, from managing customer service to helping run supply chains.

These predictive AI technologies may have found their ultimate expression at the retail giant Amazon, which deeply embraced this form of AI in the 2010s. At the heart of Amazon's logistical prowess lies its AI algorithms, silently orchestrating every stage of the supply chain. Amazon integrated AI into forecasting demand, optimizing its warehouse layouts, and delivering its goods. It also intelligently organizes and rearranges shelves based on real-time demand data, ensuring that popular products are easily accessible for quick shipping. AI also powered Amazon's Kiva robots, which transported shelves of products to warehouse workers, making the packing and shipping process more efficient. The robots themselves rely on other AI advances, including those in computer vision and automated driving.

However, these types of AI systems were not without limitations. For instance, they struggled with predicting "unknown unknowns," or situations

that humans intuitively understand but machines do not. Additionally, they had difficulty with data they had not yet encountered through supervised learning, which posed challenges to their adaptability. And, most important, most AI models were also limited in their ability to understand and generate text in a coherent and context-aware manner. Thus, while these uses of AI are still important today, they were not something most people directly saw or noticed in their daily lives.

But among the many papers on different forms of AI being published by industry and academic experts, one stood out, a paper with the catchy title "Attention Is All You Need." Published by Google researchers in 2017, this paper introduced a significant shift in the world of AI, particularly in how computers understand and process human language. This paper proposed a new architecture, called the Transformer, that could be used to help a computer better process how humans communicate. Before the Transformer, other methods were used to teach computers to understand language, but they had limitations that severely curtailed their usefulness. The Transformer solved these issues by utilizing an "attention mechanism." This technique allows the AI to concentrate on the most relevant parts of a text, making it easier for the AI to understand and work with language in a way that seemed more human.

When reading, we know that the last word we read in a sentence is not always the most important one, but machines struggled with this concept. The result was awkward-sounding sentences that were clearly computer generated. TALKING ABOUT HOW ALGORITHMS SILENTLY ORCHESTRATING EVERY ITEM is how a Markov chain generator, an early form of text generation AI, wanted to continue this paragraph. Early text generators relied on selecting words according to basic rules, rather than reading context clues, which is why the iPhone keyboard would show so many bad autocomplete suggestions. Solving the problem of understanding language was very complex, as there were many words that could be combined in many ways, making a formulaic statistical approach impossible. The attention mechanism helps solve this problem by allowing the AI model