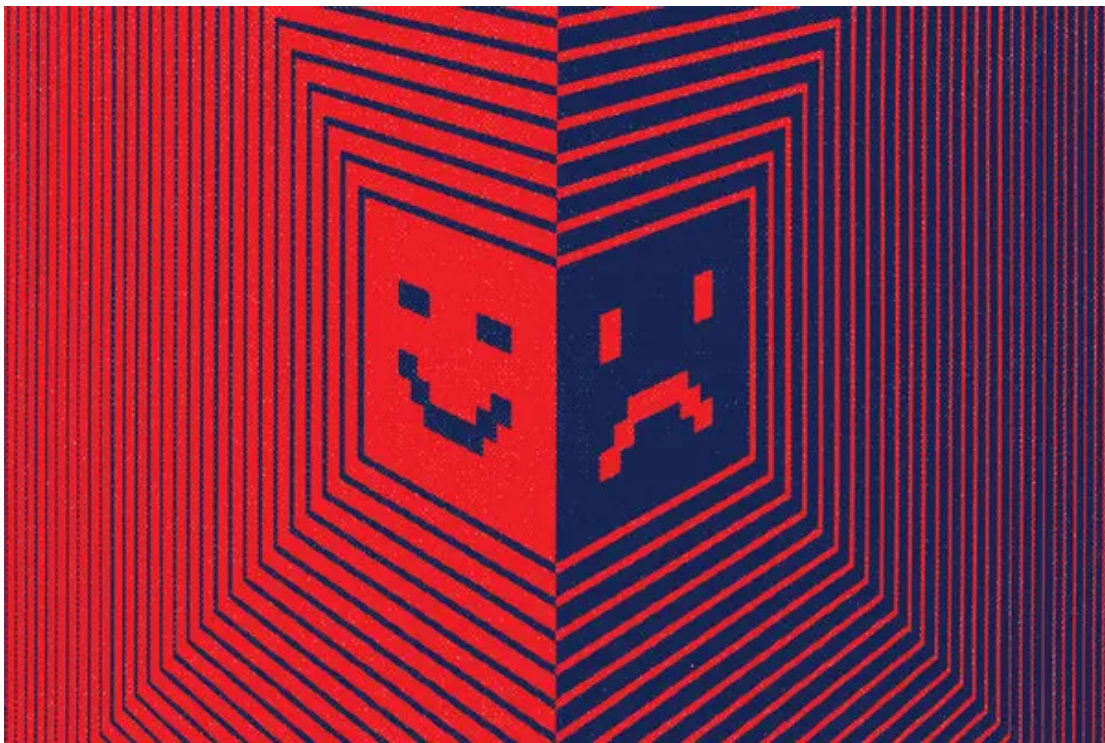


[nytimes.com](https://www.nytimes.com)

5 Best Books About Artificial Intelligence

Stephen Marche

12–16 minutes



Credit...Errol F. Richardson

Reading List

Everybody Is Talking About A.I. What the Heck Is It, Anyway?

A guide to the best books about artificial intelligence.

Credit...Errol F. Richardson

Stephen Marche

Stephen Marche is the author, most recently, of “The Last Election,” a political thriller written with Andrew Yang.

- Jan. 31, 2024

Much of the current discourse about artificial intelligence sucks. A.I. is one of those subjects that seem to appear out of nowhere, blotting out everything else — like the sun rising over the desert — and a mixture of ludicrous hype and vacuous panic has rushed in to shade the blinding dazzle. A.I. is going to transform every industry, or it’s going to end the world, or both, and you need to know which now. Meanwhile, accurate answers to basic questions — What even is it? How does it work? Where did it come from? Where is it going? — tend to be tucked away in dry technical language that borders on the incomprehensible.

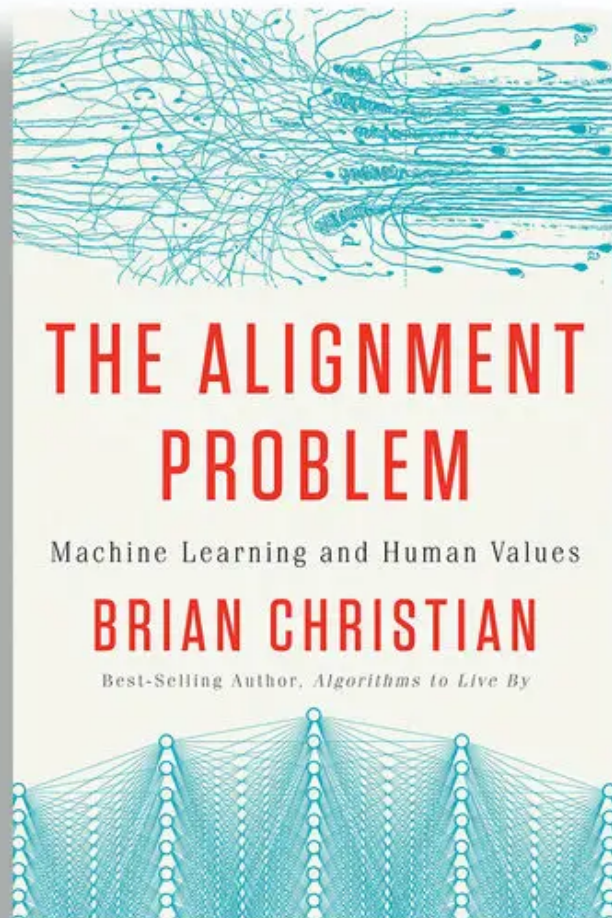
Before the [launch of ChatGPT](#), a little more than a year ago, it was difficult to get readers to care about A.I. Sam Altman, the chief executive of [OpenAI](#), would tell anybody who would listen that A.I. [needed regulation](#), but few in power *were* listening. Editors sniffed. Readers yawned.

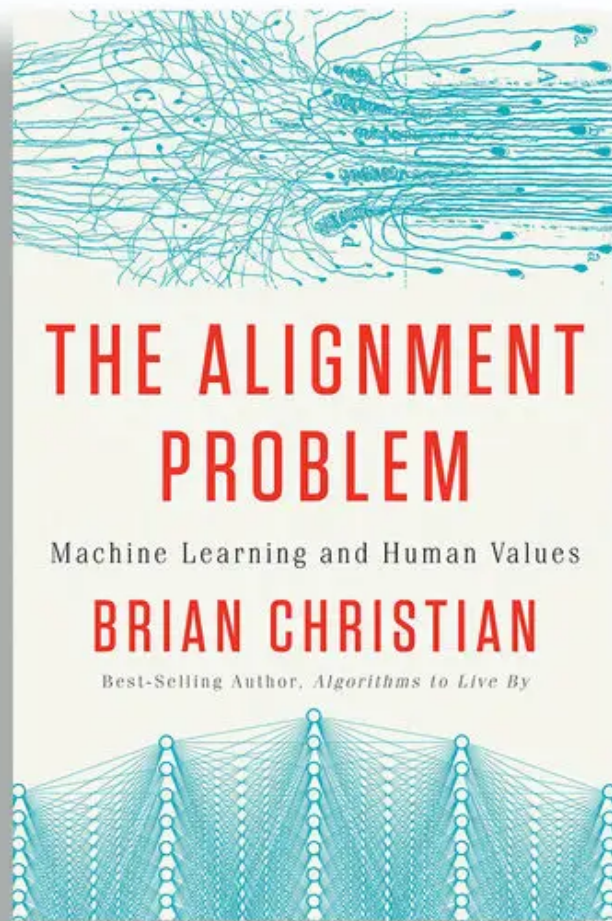
After the launch of ChatGPT, everybody had an opinion, and nobody knew what they were talking about. The novelty and the urgency provoked the usual grift that accompanies any glut of public ignorance. The movies, with their predilection for wild visions of the artificial intelligence future (A.I. will [start nuclear war](#), [enslave humanity](#) or [teach us the nature of love](#)), didn’t help. And, after a decade during which Silicon Valley has demonstrated that [it lacks any sense of social responsibility](#), it has become impossible to trust the creators of A.I. Then there is the confounding nature of the technology itself, which often eludes the understanding even of the people who invented it. It’s amazing that anything good about A.I. ever gets written.

Such books tend to come in two principal flavors: “We’re all going to

die” and “How to get rich.” You can easily judge them by their covers. Ignore them. The good news is that there *are* some terrific books about A.I. once you weed out the grifters.

Image





The Alignment Problem, by Brian Christian (2020)

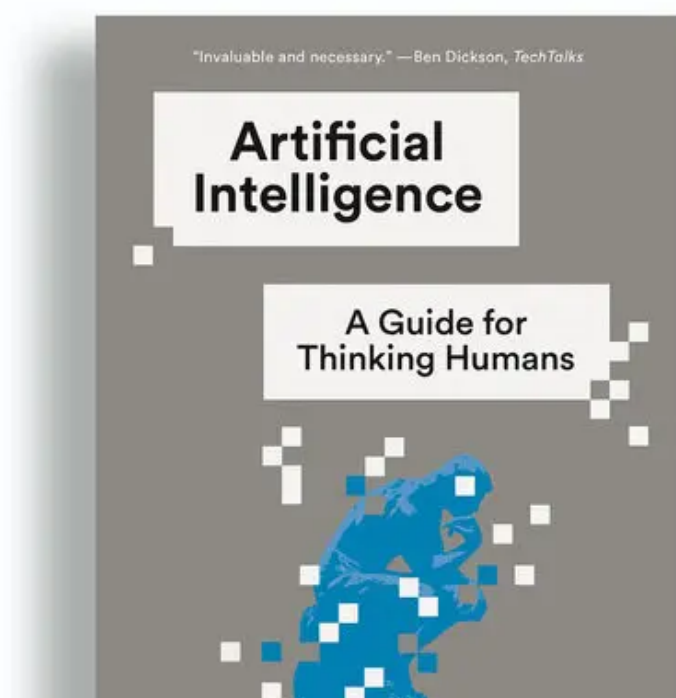
If you're going to read one book on artificial intelligence, this is the one. Though it was published in 2020, which in terms of A.I. is practically prehistory, I still think it's fairer and more illuminating than almost anything published since. Its chief value is its close examination of the computer scientists, cognitive psychologists and philosophers who were present at its birth. You just can't beat dense reporting.

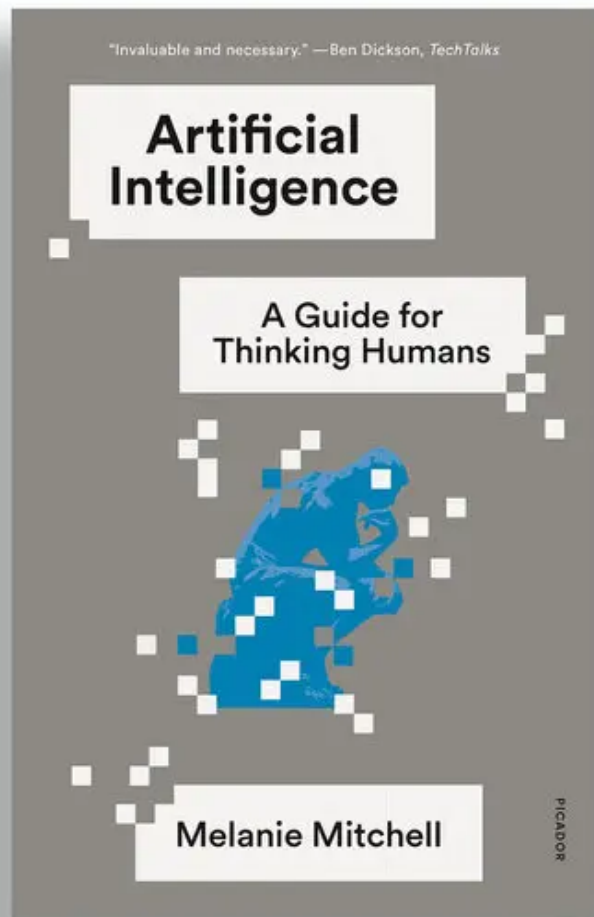
The problem with A.I. isn't that it's going to end the world, Christian says. The problem is determining how to "align" machine behavior with human values, a conundrum we have been trying and mostly

failing to solve since the invention of the cotton gin. “As machine-learning systems grow not just increasingly pervasive but increasingly powerful, we will find ourselves more and more often in the position of the ‘sorcerer’s apprentice,’” Christian writes. “We conjure a force, autonomous but totally compliant, give it a set of instructions, then scramble like mad to stop it once we realize our instructions are imprecise or incomplete — lest we get, in some clever, horrible way, precisely what we asked for.”

Christian, the author of two previous books about the intersection of humans and computers, is admirably clear: The trouble isn’t only the machines; it’s the people. To align machines with human values, we have to know what human values are, and that knowledge is hazy at best. We cling to theories of fairness and transparency, which grow vaguer and vaguer when we try to put them into practice. While Christian is hopeful — this is a book that celebrates A.I. as a victory for scientific progress and doubles as a manifesto for A.I. safety — he is also realistic. “Alignment will be messy,” he concludes. “How could it be otherwise?”

Image





Artificial Intelligence, by Melanie Mitchell (2019)

This is the Honda Civic of A.I. books; I mean that as a compliment. It's reliable, it's durable; it gets you where you need to go without a lot of fuss. Its author, a [professor now at the Santa Fe Institute](#), is a longtime specialist in computer science and complex systems; the

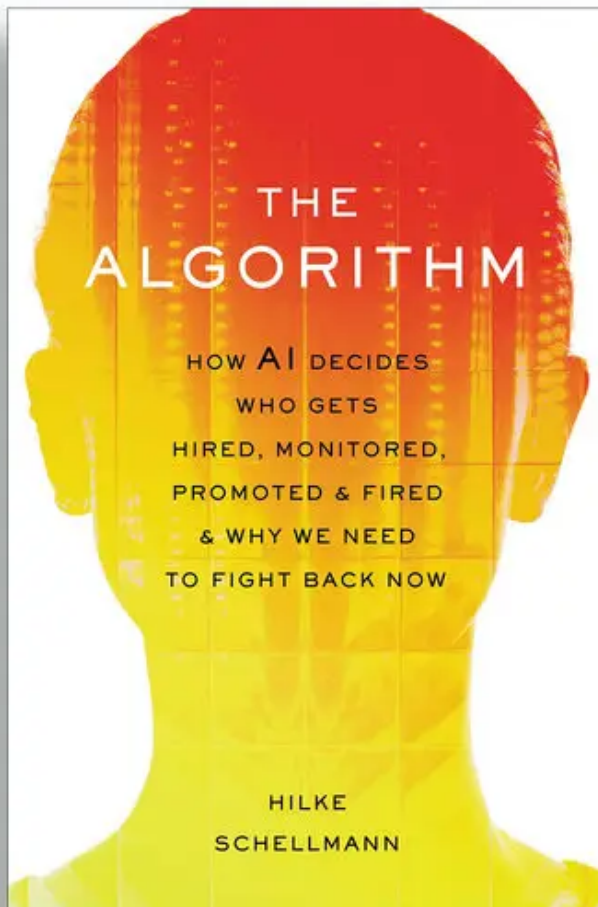
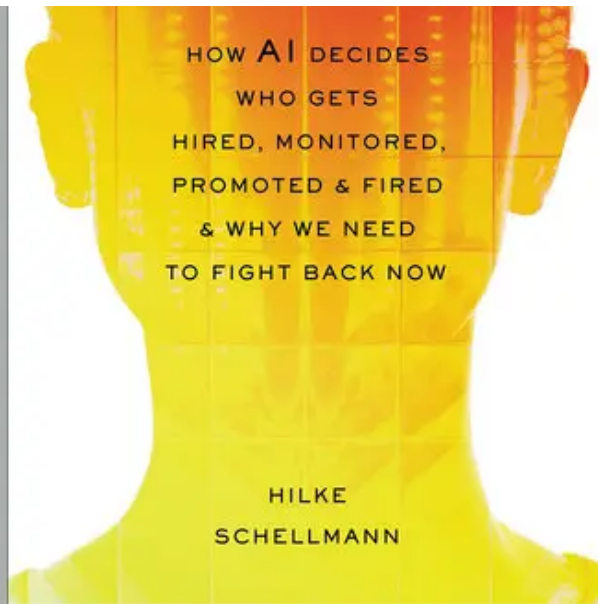
number of people who have worked in A.I. and can string together a coherent sentence is perishingly small, which is one reason her book is so useful.

“Artificial Intelligence: A Guide for Thinking Humans” provides the necessary history of the development of the technology, from the psychologist [Frank Rosenblatt and his perceptron](#) (a simple neural network) in the 1950s to the [Stanford computer scientist Fei-Fei Li](#) and her [ImageNet](#) (a landmark visual database used for training intelligent machines). The story Mitchell tells is chronological and detailed, encompassing the intellectual breakthroughs of the Dartmouth College group that coined the term artificial intelligence in the mid-1950s as well as the advent of natural processing language in the 2010s. She answers essential questions about artificial intelligence simply and elegantly.

Moreover, Mitchell doesn’t hide the confusion that any honest encounter with artificial intelligence occasions. “The field of A.I. is in turmoil,” she writes. “Either a huge amount of progress has been made, or almost none at all. Either we are within spitting distance of ‘true’ A.I., or it is centuries away. A.I. will solve all our problems, put us all out of a job, destroy the human race or cheapen our humanity.” That’s not the simplistic take you’ll find on the opinion pages, but after reading Mitchell’s guide, you’ll know what you don’t know and what other people don’t know, even though they claim to know it. And that’s invaluable.

Image





The Algorithm, by Hilke Schellmann (2024)

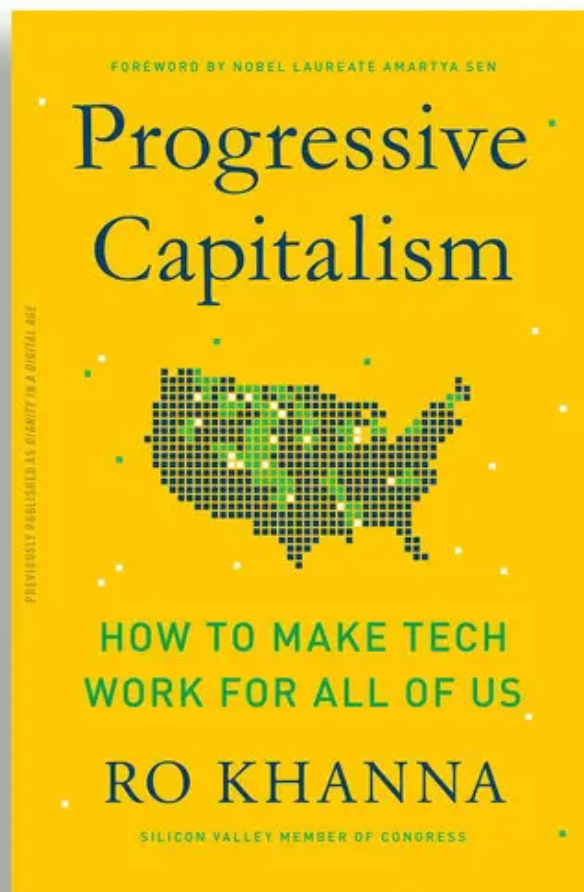
Terrible title, right? I mean, at this point a book called “The Algorithm” could be about anything. I imagine the title is so vague because the book’s subject might seem unappealing on its face. What could be drearier than a book about the use of artificial intelligence by human resource departments? But if you want to know the nitty-gritty of the alignment problem — how people are actually responding to it — “The Algorithm” is the best available case study.

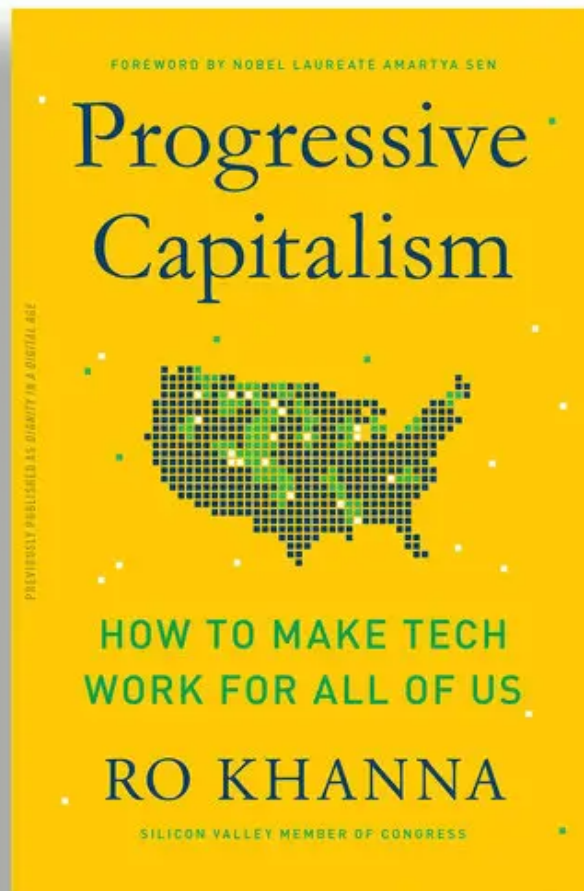
Human resources makes an excellent point of entry into the problems of artificial intelligence, because it involves so many different practices: hiring, evaluating, surveilling and retaining employees. The theoretical benefits of A.I. are obvious: increasing efficiency and reducing human bias. Equally obvious are the potential abuses. Schellmann, an investigative reporter and journalism professor, describes both, searchingly and steadily. The great strength of her book is that it treats A.I. as a tool used by people, avoiding grand theories and wild speculations in favor of close attention to detail.

She begins her investigation with a great deal of hope for this new technology, given how inept human beings are at dealing with people. “I was so excited when I started this journey researching A.I. in the world of work,” she writes. “‘Finally,’ I thought, ‘a solution to biased human hiring.’” By the end, however, her enthusiasm had waned: “At least some of the tools people and companies use to make employment decisions do not work. At least some companies are basing high-stakes employment decisions on biased and junk algorithms, which cause real harm and prevent qualified people from getting jobs.” Human beings aren’t very good at making decisions in the first place, but A.I. does not liberate us from our human limitations; it chains us to them.

The nightmares Schellmann describes — the use of facial recognition software in hiring decisions and the drive toward total employee surveillance — are more chilling than any apocalyptic scenario because they are actually happening. At times, “The Algorithm” reminded me of a digital version of Charlie Chaplin’s ordeal in “Modern Times,” with our minds rather than our bodies pulled and twisted through elaborate gears.

Image





Progressive Capitalism, by Ro Khanna (2022)

The alignment problem is fundamentally political. Any meaningful response to artificial intelligence will have to be collective. But who will the collective be? The idea that Silicon Valley can self-regulate is by now a sick joke, yet the American government doesn't seem much better positioned to do the job. Fortunately, Khanna, a Democrat representing California's 17th Congressional District, offers some hints of what a serious political response to A.I. might look like.

"Progressive Capitalism" is a book written by a politician, and therefore filled with the politician's standard scenes (visits to coal

mining towns you've never heard of) and rhetorical devices (digestible five-point plans about issues of unimaginable complexity). But if you skip over the filler, Khanna offers one of the strongest prescriptions around for how to deal with the economic and social upheaval inevitable in a transition to new technologies. "While we allowed the internet to emerge without a strong ethical or legal framework in place, we should not make the same mistake when it comes to the emergence of artificial intelligence," he writes. He cites a PricewaterhouseCoopers report that concludes that by 2030, A.I. will add over \$15 trillion to the global G.D.P., but notes that almost all that value will go to a sliver of the population. (At the moment, he adds, only 0.0006 percent of venture capital in the United States goes to Black women.)

Khanna proposes "technology hubs" in the Midwest and South to spread the benefits — by supporting research and career training — while remaining pro-growth. He is particularly strong on preventing data discrimination, potentially the most socially disruptive element of artificial intelligence. "Ultimately, there must be clear rules that impose liability on institutions that use personal data in discriminatory ways or that rely on algorithms that further disparities based on race, gender or other demographic considerations," he writes. Without such protections, algorithms will exacerbate the human failings that already pervade the tech industry.

These days, almost nobody turns to American political institutions to figure out what's going on or what to do about it. [Trust in Congress is in collapse](#), and even at the best of times its members have demonstrated a limited understanding of technological change. (Remember Senator Ted Stevens's description of the internet as "[a series of tubes](#)"?) But Khanna knows what he's talking about, and it's comforting to realize there's at least one

person in Congress who does.

Image





AI 2041, by Kai-Fu Lee and Chen Qiufan (2021)

Anyone who writes about artificial intelligence has to speculate; it's the nature of the subject. But A.I. has a way of humiliating predictions. (In 2018, Lee [predicted Chinese domination of the field](#), which the advent of ChatGPT instantly disproved.) But it's too fun not to speculate, and "AI 2041" is self-aware fun.

Lee, a Taiwanese-born, American-educated venture capitalist who is a prominent figure in China's A.I. scene and a former president of Google China, provides the intellectual foundation for the book, but he was smart enough to join forces with a proper writer, Chen, a [leading author of realistic science fiction](#), to articulate the visions of

the future. Each chapter consists of a story by Chen followed by explanations of key concepts by Lee. The approach separates fantasy from verifiable information, which serves both sides well. The nonfiction isn't lying. The fiction is backed up by facts.

The range of the speculation is impressive, too, spanning continents as well as technologies. A Nigerian video producer generates deep fakes. Korean orphans learn from cartoonish A.I. tutors. In Sri Lanka, autonomous vehicles have unintended consequences. Lee's commentary tackles whichever technological innovation has inspired the story: computer vision, natural language processing, bitcoin security, etc.

In other books, explanations of the mechanics of artificial intelligence tend to be either drearily technical or childishly reductive; "AI 2041" has found a clever way of avoiding both dangers. In recognizing that the deeper consequences of the A.I. revolution are fundamentally unknowable, it upholds [Amara's Law](#): "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run."

There is so much more to artificial intelligence than the fear and greed that have dominated the discussion so far. Artificial intelligence is magical, mysterious and profound — and not at all easy to understand. The more definitive or absolute a writer purports to be, the less you should trust their work.

These five books have the necessary humility to deal with the glamorous uncertainties. And they all uncover the strangest aspect of A.I.: Its problems are people problems. This is the ultimate irony of the new technology. The true gift of A.I. may be what it reveals about human limitations and failings, not those of machines.

Advertisement

[SKIP ADVERTISEMENT](#)

