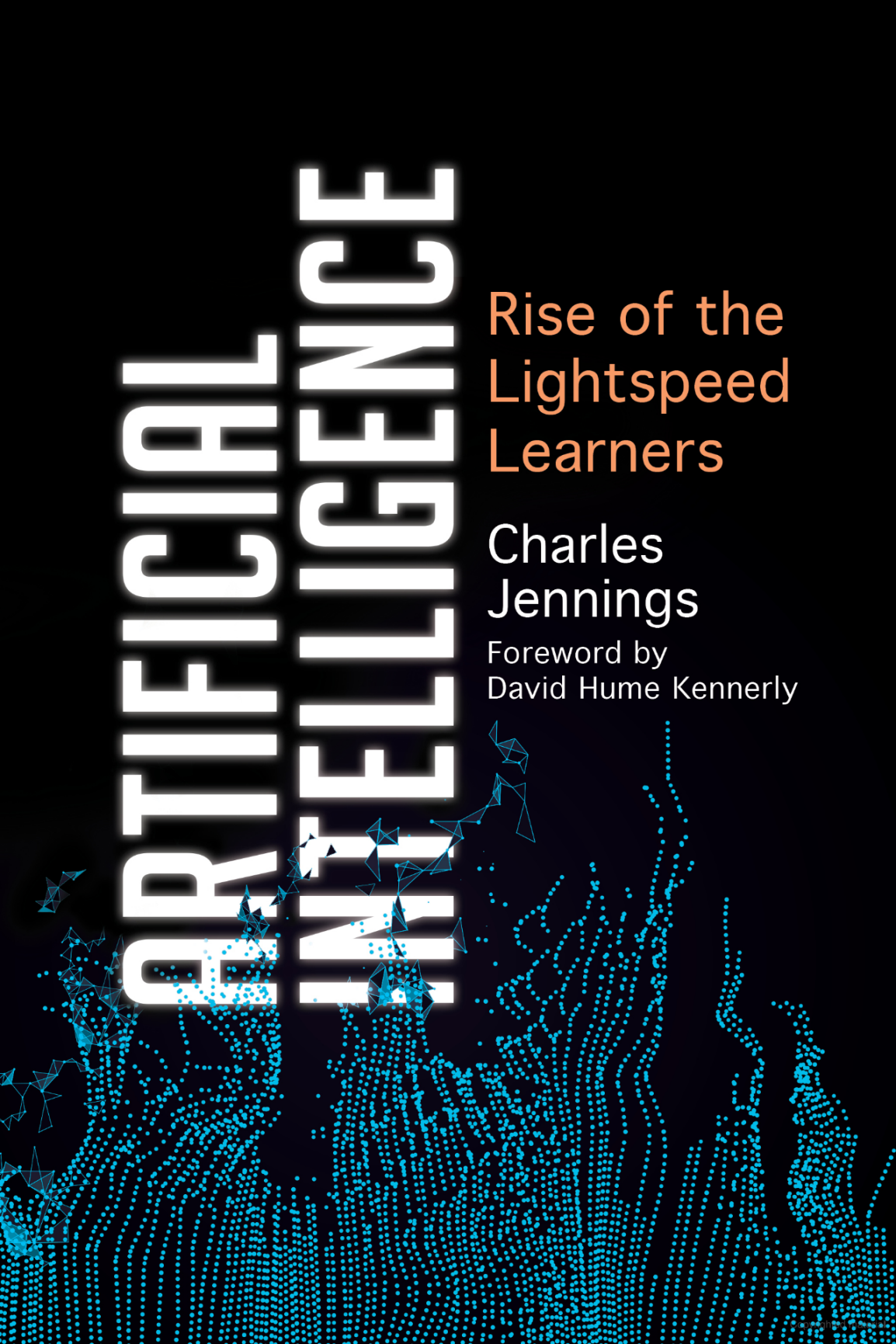
ARTIFICIAL INTELLIGENCE

Rise of the Lightspeed Learners

Charles Jennings

Foreword by
David Hume Kennerly



Artificial Intelligence

Rise of the Lightspeed Learners

Charles Jennings

ROWMAN & LITTLEFIELD
Lanham • Boulder • New York • London

Published by Rowman & Littlefield
An imprint of The Rowman & Littlefield Publishing Group, Inc.
4501 Forbes Boulevard, Suite 200, Lanham, Maryland 20706
www.rowman.com

6 Tinworth Street, London, SE11 5AL, United Kingdom

British Library Cataloguing in Publication Information Available

Library of Congress Cataloging-in-Publication Data

Names: Jennings, Charles, 1948 author.
Title: Artificial Intelligence : rise of the lightspeed learners / Charles Jennings.
Description: Lanham : Rowman & Littlefield, an imprint of The Rowman & Littlefield Publishing Group, Inc., [2019] | Includes bibliographical references and index.
Identifiers: LCCN 2018049541 (print) | LCCN 2018060314 (ebook) | ISBN 9781538116814 (electronic) | ISBN 9781538116807 (cloth : alk. paper)
Subjects: LCSH: Artificial intelligence.
Classification: LCC Q335 (ebook) | LCC Q335 .J45 2019 (print) | DDC 006.3—dc23
LC record available at https://lccn.loc.gov/2018049541

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI/NISO Z39.48-1992.

Printed in the United States of America

Contents

Foreword

Charles Jennings knows a lot about many subjects, but technology and China lead the list. This important book weaves the two together in a way that should make everyone sit up and pay attention. Like the spaceship from *The Day the Earth Stood Still*, artificial intelligence (AI) has arrived, and *Artificial Intelligence: Rise of the Lightspeed Learners* describes the moment when that alien vessel's door swings fully open.

As a veteran photojournalist, I've been an avid user of technology. In the 1960s, 1970s, and 1980s, that meant that I transmitted my photos through a telephone line, hoping that nobody interrupted the call. In the late 1990s and early 2000s, I was among the first of my generation of photographers to move from film to digital cameras. That meant a whole different way of sending pictures, and it changed the world through the speed at which news and images are delivered.

In 2013, armed only with an iPhone, I shot pictures every day for my book *David Hume Kennerly on the iPhone*. The idea was to show how quality photographs can be captured with a cell phone camera as a normal part of daily life. I've always been more interested in using advanced technology than understanding exactly how it works. In other words, I am neither a photo geek nor tech geek! What's important to me is what I see through the lens and to freeze that moment for all time in a compelling image. What I haven't been sure about is how AI will affect what I do as a professional documentarian, but with Jennings as my professor teaching this old dog some new digital tricks, I know that change is going to come.

As everyone in the news and tech business knows, something big is brewing with AI, and it has a scary edge. As author Stewart Brand said, "Once a new technology rolls over you, if you're not part of the steamroller, you're part of the road."

Charles Jennings, whom I've known and worked with for more than thirty-five years all across the planet, is a steamroller. He has been CEO of several significant Internet companies, as well as an AI company affiliated with Caltech. He believes that we all need to learn about AI, and pronto. I even detect an evangelical fervor to his mission.

Since 2014, he has been schooling me about neural networks, deep learning, Google Brain, and AI vision systems. He's particularly focused on how AI is creating the biggest technology disruption ever and how important it is that "we citizens" understand and appreciate AI. And it's not just because AI will affect our lives and our livelihoods but also because our government will soon be making crucial decisions about this powerful technology. We citizens cannot afford to remain ignorant of its ramifications.

Jennings has adapted a naturally intelligent approach to explaining AI—and it's not an easy subject. It helps that he has many personal stories to tell. Jennings and I are intrepid world travelers. We have worked together in China, Thailand, Japan, Hollywood, and Washington, DC. Of all those places, China was always our favorite. It's no surprise to me that a key focus of this book is about what is happening today in the People's Republic of China. Chairman Xi Jinping is placing a huge bet on AI, as this book makes clear. Given the power of AI technology, combined with Chinese clout, this is something that should alarm us. It is helpful to have Jennings, who has worked extensively in both Silicon Valley and Beijing, as our guide through this not-so-artificial minefield. I first traveled to China in 1972, just a few months after Richard Nixon made his historic trip. On that journey to Beijing and Shanghai, I was able to photograph Premier Zhou Enlai. Something Zhou declared resonates and underscores why we shouldn't remain ignorant about the Chinese and their motives. Zhou said, "One of the delightful things about Americans is that they have absolutely no historical memory." He may have been joking, but he has been proven correct many times. China has taken advantage of an American public that is often shockingly unfamiliar with earlier world events.

Chinese history refresher: The People's Republic of China, lest we forget, is ruled by a tough-minded Marxist Communist Party. It operates under a whole different set of values and rules than we normally do here in America. Jennings uses his background in China to imagine a decisive United States

versus China AI derby in the decade ahead, the outcome of which will affect all our individual lives and those of our children and their children.

On a day-to-day basis, Jennings's stories about AI might not be the kind that make the nightly news, but their cumulative impact could be more profound than any of the wars I have covered. *Artificial Intelligence: Rise of the Lightspeed Learners* features news you can use provided by an original Internet pioneer. Charles Jennings has written an important introductory guide to the strange and compelling new world of AI. He thinks it could very well upset the current world order and unquestionably will change our lives. He believes that AI will be bigger than the Internet, and as a former CEO of companies in both the Internet and AI sectors, he knows the turf.

This book is a colorful journey into the land of AI told by an excellent storyteller and a call to action by someone who understands disruptive technology. It is must-read literature. After you have finished, please send a copy of this book to your favorite member of Congress and anyone else whose opinion might matter as we begin to face the existential challenges AI presents.

David Hume Kennerly
White House photographer, Pulitzer Prize winner

Acknowledgments

These days, building a paper book is rather retro, but dedicated people around the world still keep cranking them out. The book publishing industry continues to capture first-class talent, up and down its ranks, and supplies us all with more and better quality books than we probably deserve.

So the first tip of my Stetson hat—a gift from my twin daughters, Faith and Nayana Jennings—goes to the U.S. book publishing industry, and to my own stalwart publisher, Rowman & Littlefield. Since 1977, publishing companies have hired me to write books for them, including some big name houses. Rowman & Littlefield is the best publisher I've worked with, hands down. Kudos especially to the well-oiled team at Rowman & Littlefield that produced this book, including: my editor Suzanne Staszak-Silva, production editor Andrew Yoder, copyeditor Niki Guinan, and cover designer Devin Watson.

Frankly, all my experiences with publishers and book agents have been pretty good. I've worked in movies, high-tech, gov-tech, and AI, but none of these industries have smarter, nicer, more honest people book publishing. Keep up the great work, all you book peddlers!

I must of course acknowledge the contribution of my good friend, Pulitzer Prize winner and super salesman, David Hume Kennerly. His foreword to this book got my wife's coveted, and rare, two-thumbs-up approval. David helped me in a variety of ways with this book: connecting me with Scott Parazynski, the MD astronaut; setting up my day with Jon Meacham; brainstorming about our time in Asia in the 1980s and about China today. In the '80s, David was an Asia-loving rock-star photog for *Time* when it ruled print media; I was a

Hollywood line producer working in Asia for Warner Bros., Paramount, and others. But as they say, what happens in Kanchanaburi stays in Kanchanaburi.

Another vital contributor to this book was my granddaughter, Ash Olsen. Raised in Oregon, she had the chutzpah to study math at Queen Mary College, London, and graduate with honors in statistical analysis. She had a few weeks downtime after graduation and before returning to America, so naturally I hired her to help me get my facts, citations, and end notes straight. She turned out to be an enormous help, at precisely the right time.

Without the work of my agent, Jeff Herman, this book would not have been possible. A tired cliche, but literally true in this case.

A special tip of the Stetson to all who directly contributed AI-related content, including notably Mat Jennings, Andrej Szenasy, David Barton, Ash Olsen, Scott Parazynski, Christine Jennings, Max Tegmark, Pete O'Dell, Jon Meacham, Eric Pulmer, Dmitri Tuzoff, and Willard Ptotsman.

Finally, I am most grateful for the continuing love and support of my fifteen fellow members of the Portland Jennings clan—especially my wife of 46 years, Christine Jennings, who has learned how to be a writer's perfect companion.

To all of you, my heartfelt thanks.

CHAPTER ONE

An Uncanny Ability to Learn

"The only thing we know about the future is that it will be different."

—Peter Drucker

After decades of academic captivity, AIs have escaped their lab cages and are swarming out into the real world. With consequences for all of us. Artificial intelligence (AI) comes in many forms, sizes, and algorithms. Today, you'll find AIs in factories, schools, hospitals, banks, police stations, and the chip in your iPhone. They're the eyes of self-driving cars, the speech of Siri and Alexa, the brains of autonomous drone warfighters. They're the wizards behind weather forecasts, the intelligence that guides robotic hands assisting in colon surgery.[1] AIs handle real-time scheduling for the multibillion-dollar vacation rental market and do the matchmaking at Match.com. The hottest new online games have AI players always available as a clever opponent, 24/7. An AI has written a Grimm's-style fairy tale (*The Princess and the Fox*), and the first AI-enabled toothbrushes clean teeth intelligently. The business plan for Elon Musk's next company, after PayPal, Tesla, SpaceX et al., is to embed AI chips in humans. For medical purposes only—at least for now.

On NASA's Mars 2020 mission, advanced AIs will autonomously pilot four rovers exploring lava tubes on Mars.[2] An AI-led mission to Alpha Centauri is planned for 2069. Astrophysicists even talk of AIs exploring virtually the entire cosmos. *2001: A Space Odyssey*'s HAL computer would be proud.

As a high-tech industry, artificial intelligence couldn't be hotter. AI tops all current venture capital (VC) investment categories and leads corporate

research and development (R&D) spending. Amazon, Microsoft, Google, Facebook, Intel, Apple, and IBM all have big internal AI groups, and each is in the same elite, clubby AI alliance.[3] On Silicon Valley résumés, AI gigs are the sexiest bullets. AI engineers are in such demand that tech giants are buying premarket AI startups just for their employees—paying up to $10 million per head—not per engineer but per *employee*, from CEO to front-desk receptionist. And this is just in the United States. If anything, China is making even more aggressive investments in AI start-ups.

AIs are flying high, but with serious baggage. For example, some very smart people believe that an AI might someday become the next Joseph Stalin. Investment bankers (Goldman Sachs) and consulting firms (Deloitte, PWC, McKinsey) are predicting that one-third of the current American workforce will lose their jobs to AIs in the decade ahead. And why are China and Russia suddenly so gung ho on AI, and what should we be doing about it?

Don't look now, but there are many fewer people these days working in auto factories, lettuce fields, stock exchanges, distribution warehouses, call centers, air-traffic-control towers, clerical desks, customer support centers, and most retail big box stores. Have you noticed? It's not that some strange force has called these workers to another world. They've being replaced by AIs.

Amazon Go—the world's first fully automated grocery store—is up and running in Seattle, with no human sales clerks. Shoppers download an Amazon Go app, hop an Uber, fill out their list in the car, arrive, gather up their groceries, and leave. Quoting an Amazon Go ad, "No lines. No checkout. Just grab and go." Not good news for retail clerks.

The trend of machines taking human jobs is not just going to continue; it's going to explode, like mortars across Kabul, causing nasty disruptions and leaving real victims. The people losing jobs this time will not only drive trucks and work in factories, but they'll also practice law, prepare tax returns, manage personal wealth portfolios, teach university classes, and even practice certain kinds of medicine. Expect to see radiologists, actuaries, commodities traders, paralegals, marketing consultants, professors, pharmacists, and more as collateral carnage (economically speaking) lying by the roadside.

An AI-driven job-market disruption is coming; the only question is *when*. If the job losses described here are spread out over two generations, no one will much notice. If they're in full swing by 2025, as most I-banks and consulting companies predict, Western economies will be rocked to the core. Robots taking over auto workers' jobs in America, slowly over two decades, was a big deal. It caused an economic disruption in the upper Midwest that helped Donald Trump get elected president of the United States. Who knows what would happen if AIs replace half of all lawyers and accountants?

Part of the problem is that Silicon Valley culture has long regarded lost jobs as wet garbage. "Move fast and break things" is its near-official motto—a sentiment that does not lead to much empathy when new tech kills a job category. I can't tell you how many times in the valley I've heard some variation of "Those bookstore owners (or travel agents or cab drivers) should have seen it coming," which provides the justification for the standard Silicon Valley response when new tech creates job losses: "Hey, deal with it." This time, with AI, the technology community must do better. This time we must all do better.

This book is less about the future moral and philosophical implications of artificial intelligence—a favorite publishing theme of late—than it is about the fact that AIs are here to stay. Multitudes of them of various kinds, swarming everywhere. AIs are part of an unprecedented disruption, the invasion of new forms of intelligence on earth. This book is largely an attempt to make sense of this invasion, in social, political, and economic terms.

Unlike most AI books, this one comes with neither neat equations nor definitive solutions to artificial intelligence issues. Rather, it is a series of stories, explorations, and questions—even a bit of humor and poetry. My task as narrator, as I see it, is not only to introduce you to this new species, to these machines that can learn with such extraordinary speed and power, but also to get you *thinking* about them, as if they really do matter, now, in your life. My goal is to encourage you to act, as a worker, as a consumer, and as a citizen, in ways that will help shape AI's future, and your own.

There is debate about the number of new jobs AI will create but almost none about the tremendous number it will kill. Over time, every task that can be routinized will be, even if it involves higher-order cognition. Predicting the timing of specific job market disruptions is always tricky, as I show in chapter 4, "Truckin' in Flip-Flops," about self-driving trucks. But the job losses will keep on coming. One interesting hit already has come to Goldman Sachs' currency trading division, where human employment is down 99 percent since 2010 and where AIs and computers now do almost all the work (see chapter 10, "The AI Casino").

Many AI experts believe jobs will not be the only things AIs take from us. AI may become, as James Barrat titled his important 2013 book on artificial intelligence, "our final invention"—not the last invention on Earth; just the final one invented solely by humans. In 2017, Google's CEO Sundar Pichai announced that his company's AutoML unit had successfully taught its machine-learning software how to program new machine-learning software on its own, and it does so better than humans do in some cases.[4] The *machines-programming-machines* era of AI development has begun and will

no doubt gain strength dramatically in the decade ahead. When it comes to invention, today's AIs are like teenagers with learning permits, teens who soon will become the young adult drivers of innovation.

But the notion of AIs someday supplanting all human innovation is, in my opinion, as unlikely as it is dreary. We humans are and will remain indispensable to the civilization we have created and built, so long as we have the will to do so and so long as we exercise a modicum of control over these new AI creatures. Still, I worry about the rise of these *mathsects*,[5] as I sometimes like to think of them. I worry that they are spinning us faster and faster toward some chaos we cannot control. But I also believe it is equally possible that AIs might emerge as a kind of technological superhero, fighting at our side for truth, justice, and the American way. I have been an entrepreneur all my life, so optimism runs in my veins.

My wife is a beekeeper, and I help her a bit. We've kept bees in our meadow alongside Cedar Mill Creek outside Portland, Oregon, for about a decade now. AIs are like smart bees that feed on ones and zeros, on data. When well-fed, they have the potential to produce the sweetest honey and most nutritious royal jelly—or sting you like a pissed-off scorpion.

I entered the AI industry (sans bee suit) in 2014. I had what in Zen is called beginner's mind; in Silicon Valley lingo, a blank whiteboard—not necessarily a compliment. The first two things I learned were:

1. AIs today are a primitive, immature species, but even so, they are learning at rates orders of magnitude faster than we humans ever have.
2. Just like bees, once AIs start swarming, no one can say for certain where they're going to land.

AIs are the most important technology of my lifetime. I say this having witnessed the rise of the Internet firsthand as the founder of two Internet companies in the 1990s. But the Internet is small change compared to having these lightspeed learners buzzing around.

Artificial intelligence is not a *thing*; it's an ingredient in everything. Or, more properly, it's a class of things that have this in common: *an uncanny ability to learn*. Hence the *lightspeed learner* designation in the title of this book, a term I coined to underscore the amazing learning capabilities of AIs.

Over the past several years, the latest generation of AIs—the ones doing odd things like deep learning, Monte Carlo tree searches, tensor processing, and so forth—have been absorbing knowledge and acquiring skills in record time. Chapter 2, "Not Your Father's AIs," explores how these latest AIs learn, but meanwhile, consider the case of DeepMind versus Go.

DeepMind is an AI research arm in Google's Alphabet soup; Go is the oldest and most popular board game in the world, invented 2,500 years ago in China. The DeepMind team built various versions of a game-playing computing system over the course of several years. The first version, AlphaGo, was supervised and trained by humans, and after six months of digesting rules, studying expert moves, and playing human opponents, it was able to beat the best Go player in Europe. Three months later, it beat a prominent Korean Go master in front of a huge audience watching on Asian television.

Then, in December 2017, the DeepMind team launched a variant called AlphaZero, which *taught itself* to play Go, as well as chess and shogi, in a couple of days.[6] It did this by digesting the rules of these games and then playing itself over and over. Without any human-supervised training, AlphaZero was able to beat AlphaGo and every other top game-playing computer in all three games. AlphaZero supassed sixty-plus years' playing of computer chess from scratch in less than a day.

The performance improvement from AlphaGo in 2016 to AlphaZero in 2017—accomplished in reduced training time with nearly complete machine autonomy—is emblematic of the progress AIs are making in many fields right now. AIs are not just learning; they are also learning at ever faster rates. This undisputed fact has sparked a furious global debate.

Bout of the Heavyweights

In one corner, such icons as Elon Musk and Bill Gates issue doomsday warnings. The late Stephen Hawking in 2014 said, "Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last." He also said, "[T]he development of full artificial intelligence could spell the end of the human race."[7] Musk has said that AI could become an "immortal dictator from which we never escape."[8] Even Sergey Brin, cofounder of Google and current president of Alphabet, Google's parent company, says, "We are on a path [with AI] that we must tread with deep responsibility, care, and humility."[9] In the other corner, Ray Kurzweil, Jeff Bezos, Mark Zuckerberg, and most rank-and-file AI engineers insist that artificial intelligence is the greatest invention since fire, it will be a tremendous boon to humanity, and there really is nothing we should worry our pretty little heads about. The one thing both camps agree on is that AIs will soon become much more powerful than they are today.[10]

The People's Republic of China (PRC), too, is suddenly swarming with AIs. Artificial intelligence has become the PRC's number 1 economic and

technological priority. In a national campaign modeled after its hugely successful high-speed rail program, the Chinese government is funding AI studies in universities, launching AI research labs, and orchestrating investments in private AI companies. It also has an important initiative to bring its AI companies into government R&D programs, including for military defense and homeland security. SenseTime, a Beijing AI computer vision start-up with several Americans in its senior ranks, received more than $500 million in PRC investment its first two years. One of China's hottest TV shows is a poker tournament where humans play against an AI program known as Old Poker Master. Old Poker Master always wins.

Rumors abound of American AI engineers getting offers of $500,000 a year or more to work in China. Chairman Xi Jinping wants and expects China to become the AI world leader. Given Xi's resolve—and the unbelievable amount of personal data the Chinese state controls without privacy constraints—I wouldn't necessarily bet against the People's Republic becoming the AI leader by 2030, which is Xi Jinping's oft-stated goal (see chapter 7, "Uncle Sam versus Red Star"). Meanwhile, Vladimir Putin keeps shouting, "Whoever leads in AI will rule the world!" Russia is not a major AI player, except in defense, space, and election hacking.

Today, North America is AI's clear epicenter. Of the top one hundred AI experts in the world, ninety-five are citizens of the United States or Canada.[11] The US economy still spends far more on AI research than any other, and American-owned companies continue to set all the global performance benchmarks.

France, under the leadership of Field Prize–winning mathematician Cedric Villani (known as the Lady Gaga of math for his eccentric attire), has recently embarked on a compelling—if as yet underfunded—new national AI strategy. This strategy, called AI for Humanity, could become a model of government policy for AI development and governance around the world.

Militarily, the United States leads, but China is gaining on us, with both an aggressive government commitment to AI and robust support from its commercial high-tech companies. Russia, with little commercial high tech, does have excellent math education and is committed to huge military spending for AI research and development. Israel, France, the United Kingdom, and ninety other countries have some form of military AI testing and training underway. The plan for most of these armies is to relegate the most dangerous missions to AIs and robots. It's still early in what will almost certainly become a full-fledged AI arms race, but we can soon expect to see autonomous drones executing sophisticated OODA loops against each other

in military engagements without a human in sight.[12] Hopefully, this will be just in joint military exercises, but perhaps in actual combat.

Today's AIs can see, speak, learn, and (with some robots) think on their feet. Stunning AI performance breakthroughs are reported monthly. As I write, the United States—through the efforts of such stalwarts as DARPA, MIT, Caltech, and IBM, along with new kids like Google, Amazon, and Facebook—is responsible for most breakthroughs. Canada is the center of much progress in the powerful deep-learning wing of AI, and Cambridge, England, home of Google's DeepMind, would also get a star on any AI world map.

However, unlike France, China, and Russia and despite its continued dominance in the field, the United States has no national AI policy and no AI laws or regulations (outside of vertical domain rules, such as for flying drones). The fact that the United States still, late into the second decade of the twenty-first century, has no national plan for AI is both remarkable and negligent. Perhaps neither the leaders of American government nor most of the people who elect them are aware that a highly disruptive AI storm is headed their way. And they certainly must not be considering what it would mean if the United States were no longer the dominant force behind this storm.

Writing in the 1990s, British author David Ellis observed the beginnings of an epic battle between man and machine, with the latter emerging as a new intelligent species, one that would eventually compete with humans for what Ellis called the "stewardship of Earth." He dubbed this new species *Machina sapiens*, the thinking machine.[13] Twenty years later, when searching for terms that give these new kinds of machines their due, *Machina sapiens* still works.

A century from now, the upcoming 2020s will likely be known as the period in history when *Machina sapiens* gained a foothold on planet Earth. No matter that we humans gave rise to this invasion, the important point is that somehow, out of the global noosphere, a new species arose with an intelligence to rival our own.[14] This species is already beating us at our own games (chess, Jeopardy, and Go); managing our most sophisticated global financial exchanges; flying autonomously through the air; and doing amazing backflips on land.[15]

One of this species, a three-hundred-pound, cone-shaped robot on wheels working as a security guard at a Georgetown shopping mall, recently committed suicide, if you believe the social media meme. What's indisputable is that a robot mall cop rolled into a fountain and "drowned"—*drowning* in this case being a synonym for "shorted out its electrical circuits."

Of course, this robot did not take its own life. Actually, it is important we not fall into the Hollywood trap of anthropomorphizing AIs. AIs are not evil monsters (*Terminator*), subtle lovers (*Her*), or cuddly garbage collectors (*Wall-E*). They are—so far—idiot savants with a real talent for crunching data. Multiplying two eight-figure numbers in their "heads" takes a nanosecond; solving a quadrennial equation takes a tad longer. What matters is not that AIs can do math; it's what they can do with it. They use mathematics—the formulas in their little "algo" heads—to *learn*.

I knew none of this back in 2014, when I signed up as CEO of a new AI company. The idea for the company came from Caltech and the Jet Propulsion Lab (JPL). These actual rocket scientists had some AI patents and a bit of experimental software left over from the Mars rover program and were looking for an "engineer/executive" who could commercialize this tech and take it to market. Instead, they found me.

I'm no engineer. I do not have a PhD in math, physics, or computer science—not even close. Aside from reading Arthur C. Clarke, William Gibson, and Ray Kurzweil, I knew nothing about AI, but I've founded and led a few software companies and have written several books on technology. Somehow, I passed muster with Caltech, and we got into the AI business together.

I was not looking for a job. I was living my version of the good life in my hometown of Portland, Oregon. I was doing a little consulting; serving on a few boards; helping my wife with the garden and her bees; and teaching my tall, stocky grandson how to use his butt to rebound like Charles Barkley. Then a good friend from Los Angeles called.

"I need your help," he said. He explained that he was helping a Caltech professor commercialize AI computer vision intellectual property (IP) developed at JPL. The AI part intrigued me. Ultimately, I jumped back into the saddle of another high-tech start-up, this time, for the first time, in the emerging AI industry. I had to learn as much as I could as quickly as possible about the science and technology of artificial intelligence. Fortunately, I had good teachers.

During my AI studies, I discovered that AIs can become immensely powerful, even powerful enough (theoretically) to destroy Earth—and not just Earth, by the way, but also the entire universe. In a famous thought experiment, a superintelligent AI is programmed to optimize the production of paper clips.[16] Its raison d'être is to make more and more paper clips by any and all means necessary. Because it is superintelligent, this AI understands chemistry, biology, physics, finance, and human behavior. It knows how to accumulate money and uses the new wealth it acquires to build ever more

paper clip factories. Eventually, this AI realizes that atoms of all kinds can be turned into paper clips. Spoiler alert: The experiment doesn't end well for the universe.

My main tasks as CEO of this AI company were to raise money and recruit a team to build a practical image-recognition platform based on four JPL artificial intelligence patents. The methodology of these patents was a type of computer vision that emulated the jerky saccadic sight we humans use. One key benefit of this method is that it enables pattern recognition without laborious training and huge data training sets. As such, our tech was a part of cutting-edge AI methodology called "unsupervised learning." A number of experts, including notably Yann LeCun at Facebook, feel that unsupervised learning is the future of the commercial AI market.[17] In the end, unsupervised learning provided a compelling vision, but it proved a difficult assignment technologically, especially for a start-up competing with Google and the like.

While attempting to commercialize the Caltech IP, I learned that AIs are smart but far from perfect. The algorithms that provide the basis of AI surprised me, both with their brilliance and their stupidity. After two decades working in the enterprise software industry, I was absolutely shocked at their small size. The JPL algorithm we used for face recognition consisted of only 1,000 lines of code! The hero application at my last software company had *400,000 lines*. Yet, the Caltech algorithm was clearly more intelligent.

I learned that with artificial neural networks (ANNs)—the most common platform for machine learning—what matters is not the number of lines of code but the quality of neural operations, such as curating data effectively, then feeding these data to the AI, and conducting statistical analysis on the results; implementing feed-forward and feedback loops; and tweaking an ANN in the way a NASCAR mechanic might, dozens of times, before a big race.

AI isn't just the next PC, Internet, smartphone, or cloud. It's all these rolled together and then some—the mother of all tech disruptions.

I like to think of the difference between software and AI this way: In traditional enterprise software development, you design a blueprint; write code; and, except for whatever bugs are found, safely predict the results. With machine learning, the first result of the development process is the ability to start testing new hypotheses for process improvement. Even with highly experienced AI researchers, approximately 90 percent of their hypotheses

fail to create any improvement. Sometimes improvement comes more by accident than design. In traditional software, development proceeds until reaching a finish line called "code freeze." In our AI lab, the testing and quality analysis (QA) process never stopped. I got the feeling that ANNs were not so much programmed as organically grown. And in a very real sense, the most common outputs of an ANN are surprises.

As I got to know the AI industry better over several years, certain other things became clear:

1. **AI is accelerating rapidly.** AIs are prime examples of the law of accelerating returns, popularized by AI impresario Ray Kurzweil, which states: Not only is technology changing quickly, the *rate* of technology change is also accelerating. Yesterday's powerful new tech is being used to build tomorrow's even more powerful new tech. The cherry on top of all this acceleration is AI. With deep learning and other new machine learning methods, powerful FPGA[18] semiconductors designed especially for AI, cloud data centers offering extraordinary parallelization and scalability, and sensors collecting more data in a massively connected Internet of Things, we are now entering a perfect AI storm. Technology change is ever faster, racing like Usain Bolt, and meanwhile, the techno-geek financial industrial complex, from Google to GE to Goldman Sachs, is in full hype-cycle mode, heralding the imminent arrival of the biggest tech boom in history. And it's all just getting started.
2. **No one knows exactly where AI is headed.** The AI community is surprisingly and refreshingly open and collaborative. AI experts agree on a great deal, notably that current AI is far from the "general intelligence" we humans have. But there is much internal debate, as well. How will humans and AIs work together in future? Will AI be a job destroyer or a job creator? Will we reach the tipping point, called the technological singularity, when machines gain human-style intelligence? And if so, when? Could AIs become an existential threat to humanity? There are no consensus answers to these questions among scientists and engineers.
3. **We can all agree AI will be huge.** Except to call AI huge is to miss the point. AI isn't just the next PC, Internet, smartphone, or cloud. It's all these rolled together and then some—the mother of all tech disruptions. In recent human history, the closest things to AI were the discovery of electricity and the subsequent electrification of America in the late nineteenth and early twentieth centuries. The discovery of

nuclear energy and development of nuclear weapons and nuclear energy comes close, but ultimately, nuclear is a much narrower technology than AI and, with any luck, one whose role on Earth will remain far less significant.

4. **AI will affect all of us.** You, me, and Bobby McGee, billions of Chinese, restaurant owners in France, teenagers in India and Africa, elderly in rest homes, children in day care, women executives in Shanghai, transgender bloggers in Chicago, heads of state, kick-ass surgeons, social network divas, you who are reading this book, and I who am writing it. By 2030, AIs will be like ants at a summer picnic the day the honey spilled. They will be everywhere, affecting everyone.

There are real risks in all this. I examine the most prominent of these risks throughout the course of this book. One risk seldom discussed is that of leaving all ethical and public policy decisions about AI to my friends and colleagues in the high-tech industry, to us techies.

AI is a unique and powerful force, and we techies love unique and powerful forces. *May the force be with you.* This love of cool tech warps our vision, which is why AI needs the "force" of American democracy and culture to be programmed into its algorithms, optimization protocols, and reward functions—but that's getting ahead of our story.

In the 1990s, I started two Internet security companies and got a chance to study encryption, online privacy, and cybersecurity from inside the business world. It became clear to me that, unless security began to be designed into Internet systems at the development stage, our entire IT infrastructure would become highly vulnerable to cyberattacks. I cowrote a book based on this theory, published in 2000.[19] In it, much space was devoted to giving consumers detailed instructions on how to protect themselves from cyberattacks and identity theft. The rest of the book was a call for high-tech industry and government leaders to build more privacy and security assurance into tech products and services. In this latter effort, my coauthor and I were spectacularly unsuccessful. There are *still* not enough security controls being built into network systems, nearly twenty years later.[20]

Cybersecurity is a pernicious problem today because twenty-five years ago, when we were building all our exciting Internet and enterprise IT systems, safety was an afterthought. Security controls were either bolted on after installation or patched in after a breach. We knew enough, technologically, in the late 1990s to have greatly reduced the cyberthreat to our systems in the future, but security never became a priority—not really.[21] This failure to build cybersecurity protection and enforceable opt-in privacy policies into

our core Internet and IT systems has led to enormous data losses now, twenty years later.

The decade from 2020 to 2030 will be to AI safety and security what the 1990s were to cybersecurity. Can we, this time, get it right? Will we prioritize safety and security from the beginning or just attempt once again to install chains on the gates once the AIs are out of the barn? Will we insist that *Machina sapiens* have human-style ethics? Can we keep even cybersecure AIs from running amok, out of control? Big questions we all must ask—and ones we absolutely cannot leave merely to scientists, engineers, venture capitalists, and CEOs.

Though the rise of the Internet in the 1990s in some ways parallels the rise of AI today, this twenty-first-century AI revolution is unlike any previous technology upheaval. For the first time, we humans are not the only ones building and operating the cool new tools. The fact that machines, at least some of the time, will learn, work, and reproduce on their own changes everything.

The timing of this latest tech disruption is hardly ideal. Global warming, North Korea, Brexit, stateless refugees, the ongoing Russian hack-a-thon, the opioid crisis, global terrorism, white supremacy, species extinctions, identity theft, and a score of other first-class problems compete daily for the attention of fair-minded believers in science and human progress. In light of the great many global problems we face, managing machine intelligence can seem well down the priority list. Yet AIs could become crucial new tools in confronting climate change, diagnosing chronic disease, and solving a myriad of other problems. They could also become dangerous weapons in the hands of rogue states and terrorists. In several worst-case scenarios, AI machines themselves could become apex predators and eliminate humans altogether, which is why it is so crucial that the general public—especially those who still believe in science and respect facts—learn as much as possible about AI as quickly as possible.

In the decade ahead, we all will be riding in AI-driven cars, visiting AI doctors, talking to AI sales reps, and negotiating mortgages with AI bankers. We'll be educating our children and ourselves in classes taught by expert AIs. Career decisions will be shaped by the giant sucking sound of AI automation replacing human jobs. Small businesses will obtain a decisive competitive edge by being AI savvy—or fail because a competitor mastered AI first. We'll be choosing between different AI information services and will want to know a great deal about the privacy and security implications of each (see chapter 8, "The Porn Star's Deepfake and Other Security Paradoxes").

AIs will find cures for more types of cancers and routinely enable paraplegics to walk with the aid of exoskeletal robots. AIs and their holographic and augmented reality friends will create new immersive worlds of sound and vision. AIs will be trading stocks and managing supersmart, industry-specific cryptocurrencies on the blockchain. They will even be settling factual disputes in Congress and in courts and perhaps play major new roles in democratic governments.

Technology, especially AI technology, is racing ahead of laws, social norms, school curricula, and the comprehension of the great majority of people on earth. This is healthy neither for the high-tech industry nor for the general population. AIs are speeding downhill ahead of their skis, and the black diamond runs are just beginning.

Ready or not, AIs are invading our world. As a result of this invasion, a host of new social, economic, and ethical questions are finding their way onto center stage in modern life, including:

- What can Western democracies do to prevent a global AI arms race?
- How do we prevent the subversion of journalism by extremists using AIs to create deepfakes and other patently false "news" stories?[22]
- How much should AIs know about us?
- What are the consequences if an autonomous AI breaks the law—and who pays?
- Do we need an AI regulatory commission, of the kind established by President Truman to manage nuclear weapons?
- How do we manage AI today so that it doesn't get out of control in the future?

These are not easy questions. I certainly don't have all the answers. I'm not in the Elon Musk, Stephen Hawking, AIs-could-kill-us-all camp, at least not yet, and I believe AIs can still become powerful tools for good. But I am absolutely convinced the only way AI can become a boon and not some deadly I-bomb is if we start working together, all of us, on the challenge of safely integrating this powerful new technology into our society and our lives.

The truth is, the engineers building this stuff, the smartest of them, don't want the responsibility of making social and political decisions about AI on their own. As one said to me at an AI conference, "*Citizens* are our most important demographic." It is crucial that the general public learn about AI and become familiar with the pressing and sometimes troubling issues AI is raising, which is why I wrote this book, and with a sense of urgency. It's also

why my grandson will have to wait until next basketball season to learn the art of the Karl Malone elbow.

We in America must either engage together and control AI or watch as the Chinese—or out-of-control machines—do it for us. And we must engage in the old-fashioned way: as citizens in a democracy, working together, with government in charge.

This book has been written with the United States of America as a focus for two reasons. First, what happens in America will have great impact on what happens with AI everywhere. Second, as an American who has spent years living and traveling outside America, I have the expat's love of the homeland. Hard as it has been recently, I remain optimistic about America and convinced that the United States can—even must—play an essential role in the ongoing AI invasion.

Once Americans understand that a new national policy for AI is essential to preserving jobs, continuing economic prosperity, and saving our collective human asses, we will again raise the flag and make it clear that getting AIs under better national management is essential for national security. I fully expect an AI-focused political movement will follow—perhaps a modest one, maybe something larger. Maybe this movement can even use the forces of science and technology to create friendly political AIs—AIs that bring the power of unbiased truth seeking to our political commons, to strengthen *we the people* and reclarify our national purpose (see chapter 9, "AIs in the Government Henhouse").

Regardless of what happens nationally or globally, we all must go on living our lives, lives that increasingly will have AIs—those pesky, brilliant little mathsects—popping up like fireflies in an Ozark summer. At the end of this book, I suggest a few specific strategies for survival in the age of AI. Some of these are personal strategies; others are for businesses, nonprofits, and political groups. My objective in writing these strategies is to get you thinking about how to put AIs to work in your life—without having to learn how to write machine learning code.

Political and strategic suggestions aside, this book is mostly about my journey into the mysterious world of artificial intelligence and my reflections on what I discovered. The writer in me hopes you find it a good read.

As I said, I'm no AI expert, just a concerned citizen who has seen AI up close and who, as a result, hopes that my stories and insights can make a

contribution to the great global AI debate. My bias is toward the American government playing a major role in keeping AI safe and humane—starting with state and city governments, not federal agencies (see chapter 12, "The Way Forward"). Perhaps we can even put AIs to work for us in the urgent need to remake America herself, in the spirit of liberty and justice for all, without regard for race, color, creed, or algorithmic orientation.

CHAPTER TWO

Not Your Father's AI

> "In the game of life . . . there are three players at the table: human beings, nature and machines. I am firmly on the side of nature. But nature, I suspect, is on the side of the machines."[1]
>
> —George Dyson, *Darwin Among the Machines*

Blockchain, augmented reality, 3-D printing, nanotech, Internet of Things, apps of every description, AIs exploding everywhere. The constant parade of twenty-first-century information technologies can be daunting—dizzying, even. In every corner of the modern world, the pace of technological change is accelerating. Data are produced and stored in numbers reaching the quadrillions (1 quadrillion bytes = 1 petrabyte). Tech-driven markets of every kind are spinning faster than a PR engine on election night. Only one thing is certain: *The rate of technological change will never be this slow again.* Think about that for a moment. Discouraging, isn't it?

As a civilization, we've been pouring money and manpower into digital systems for nearly a century, and our investment is paying off. The GAFA dudes (Google, Amazon, Facebook, Apple) have driven their market valuations to unprecedented levels and are doubling down on AI investments in hopes of riding yet another giant tech wave. Microsoft, Comcast, AT&T, Verizon, IBM, and others have seats at the AI table, playing with healthy piles of chips. Huge infrastructure investment continues at such semiconductor companies as Intel, Samsung, and NVIDIA, where most of the focus these days is to prepare for the era of AI and blockchain. Meanwhile,

Internet of Things sensors, online commerce sites, and social networks are exploding as well. Global clouds connect all this constant *activation energy* (to borrow a neural networking term) and provide computing power the way utility companies supply electricity.

For decades, we've invested heavily in a massive digital technology accelerator—and guess what? It's *really* accelerating, careening ahead in ways no one fully understands. It is as if technology itself has become a self-driving car that runs faster with each passing mile. Today's intelligent machines are not sentient, and outside of whatever narrow specialization they have been trained to excel in, they are not even all that smart. But they are *lightspeed learners*—machines capable of getting smarter and smarter, with limits that are as yet unknown—machines capable, even, of rewriting their own code. Of remaking themselves.

While this does not mean we're racing inevitably toward some dystopian hell filled with killer robots, AI spies, and heartless cyborgs, it does suggest that a bit of caution is in order. Unless you'd actually prefer to be the chopped liver, you need to become a lightspeed learner yourself, at least about AI. Your financial worth may well depend on it. Honestly, your freedom and your life may depend on it because we have never before dealt with technology this powerful or mysterious.

"Who Are These Guys?"

Andrej Szenasy, head of neural operations for an AI start-up, has an hour to kill while waiting for his wife to pick him up from work. He's spent the day analyzing statistics related to the performance of his company's new face-recognition (FR) algorithm, and he's ready for a break. As Andrej will tell you, all FR systems consist of three parts: a large database of stored images ("the gallery"), new images coming into the system for recognition ("the probes"), and the artificial intelligence algorithm linking the right probes with the right gallery image.

In the old days, around 2010, face recognition was all dots, lines, and vectors. It relied on measurements between the eyes, relative position of the ears, and so forth. These early vector-based versions worked pretty well if conditions were perfect, such as when both probe and gallery pictures were taken at a well-lit, highly controlled DMV photo station. They did not work so well in a dark alley at night or when the probe was a photo of a man wearing dark glasses and a hat or a thousand other real-world situations. *Edge cases*, facial recognition experts called them.

Duplex system has already fooled people on the phone when ordering pizza and setting up hairdressing appointments.

After Turing's pioneering conceptual work and John von Neumann's critical development of a practical digital computing architecture in 1945, AI progressed in fits and starts over succeeding decades. The mid-1950s were a period of important early advances, as were the early 1980s. Then, in the late 1980s and early 1990s, artificial intelligence entered a period known as the AI winter. Research funding dried up, and little significant progress was made. For years, AI seemed to be taking two steps forward and one step back.

Why? In part because the "expert systems" approach taken by AI pioneers in the 1970s, and later championed prominently by Marvin Minsky of MIT in the early 1980s, ultimately reached a dead end. The idea behind this tack was to observe human experts—such as a chemist or architect—and then replicate this expert's work process in an AI computer program. Expert systems programmers would interview professionals to learn their rules and norms and subsequently build if/then rule sets designed to solve problems the way human experts would. These early expert systems consisted primarily of a knowledge base (extracted from the minds and practices of the experts) and an inference engine, which interpreted facts in order to solve a problem or predict what would happen if a particular decision were made.

While expert systems were adopted in a variety of specific industrial applications (notably in the oil and gas industry) and did advance the science of AI in certain important ways, they did not take AI into the realm of self-learning. They were also highly expensive and prone to errors. Ultimately, the idea of replicating logical human thought didn't work. The world is not entirely logical. To be useful, AI would have to respond intelligently to changing real-world situations and environments. Thinking machines needed to be able to think on their feet, as it were. And to do this, AIs needed to be modeled not on *what* humans are thinking but on *how* humans think biologically.

Around 2005, a "bio-inspired" approach to AI began to emerge. This new school modeled the human brain and nervous system, using a technology called neural networks that dated back to the early 1960s. Because of research funding priorities at the time, this bio-inspired movement in AI focused on human speech and vision. Natural language and image recognition systems became key AI drivers. Somewhat ironically, this emphasis on achieving human-like perception (rather than logical thought) led to new methods and architectures that also proved very good at *learning*.

The computational core of all bio-inspired AI systems is the artificial neural network (ANN). Unlike rules-based AI programming models, such as those advocated by the expert system advocates, ANNs excel at recognizing patterns and extracting key identifying features from them in order to make sense of what is being heard or seen in the real world. Lines, dots, and vectors are replaced by "regions of interest" and "unique identifiers." Using the fuzzier logic of an ANN, a scar on the face of a probe photo subject could become a key identifier. In old vector-based FR systems, it would have barely registered.

Neurons are the basic units of an ANN—roughly analogous to the nerve cells in our brains that we humans use to see, hear, think, and respond. In both ANNs and human brains, neurons "spike" when new data come in. In artificial neural networks, neuron spikes lead to interconnections with other spiking neurons. The ability to make these connections—and vary the strength of them—gives ANNs the power to create new patterns and store them, so they can be used later to recognize similar patterns. In our brain and in AIs, the image of a cat resides as a series of such patterns, all connected and all ready to be used when new visual data of a cat enters the human visual pathway. The same thing happens when a cat image is an input to an ANN.

One of the most important axioms of neural network theory is "neurons that fire together wire together." In other words, when neurons fire (or spike) at the same time, they connect in interesting and important ways. This ability of neurons to "wire together" dynamically to create patterns and pathways of understanding occupies a central role in cognition—for both humans and machines.

That's enough neural theory for now. Here's what's important for our purposes: Because of their ability to respond intelligently and at least somewhat flexibly to random sensory inputs, neural nets modeled on *biological* systems proved to be much more adaptive and reliable than any previous AI methods.

The two most prominent leaders of the bio-inspired AI school are Geoffrey Hinton and Yann LeCun. Hinton is a soft-spoken Canadian with a bad back who never sits and doesn't use airplanes; LeCun is his avuncular French-born colleague. Both are first-rate scientists and in fact worked in the same lab in the late 1980s. In the 2000s, LeCun (at Bell Labs) and Hinton (at University of Toronto) found themselves working along similar research lines and began to collaborate again. Both specialized in image recognition systems, but each was a general AI theorist, as well. Together, they perfected an ANN model called deep learning, which uses an innovative layered approach to computation. Deep learning provided a new way to hook AI fire

hoses up to big data hydrants, using a brute force computing approach to the problem of machine learning. It relied less on emulating models of human logic (as did expert systems) and more on emulating the way our human senses process information. The deep-learning model produced remarkable results and became a kind of growth hormone for the AI industry as a whole.

From 2013 until early 2018, LeCun was head of AI at Facebook. Hinton in recent years has been a leader of AI at Google. My sense is they now have a friendly but spirited rivalry. Hinton and LeCun have their jobs because their deep-learning model demonstrated a remarkable ability to process real-world inputs accurately in certain situations. Both Google and Facebook have phenomenal image-recognition capabilities—not yet perfect by human standards but very, very good nonetheless. Much of the progress in face recognition at each company is rooted in the bio-inspired deep-learning model pioneered by these two men and widely adopted by the AI industry.

When putting real-world systems into production, LeCun and Hinton relied primarily on a machine training methodology known as supervised learning, where humans "train" AIs to perform more effectively over time. This teaching or coaching consists largely of feeding neural networks large volumes of well-labeled data (this photo = cat; this photo = dog) in order to help the AI perform a specific task (recognizing dogs and cats). Supervised learning, in the context of today's massively connected data-cloud environments, has proven to be an extremely effective way to help machines learn. But as even LeCun and Hinton acknowledge, supervised learning has it limits.

"Unsupervised learning is the future of AI," LeCun has stated publicly on several occasions.[2] In unsupervised methods, there is no human oversight, no force-fed programming. Neural nets essentially learn by processing unlabeled data in much the same way human children do. Obviously, not having to go through human-intensive data labeling and supervised training gives this method a clear advantage in dynamic, unstructured situations. This ability of AIs of various types to learn autonomously and on their own is the reason that AIs are now entering a profoundly important new phase.

Defining AI

Artificial intelligence, circa 2017, is not your father's AI. It's not HAL from *2001: A Space Odyssey*, 3CPO from *Star Wars*, the nameless AI serpent in *Alien*, or the evil Skynet from *The Terminator*. And it's not just robots on the factory floor, *Jeopardy*-playing software, or self-driving cars. It's something new, something unimagined.

It is not altogether inaccurate to think of AI as a new species. A favorite term of mine for this species, as I mention in chapter 1, is *Machina sapiens*. Whatever it is called, a new life-form has suddenly started spreading across planet Earth like kudzu in the Carolinas.

Machina sapiens has long since passed the tests used by NASA to determine if life exists on other planets. It will soon be meeting Ernst Mayr's textbook definition of a biological species: a group of organisms that mate with each other and reproduce similar offspring. Algorithms will be hooking up in hard drives and other dark places every day, begetting all sorts of squirmy new algorithms.

As of this writing, artificial intelligence is still a human tool. Yes, machine self-learning is popping up everywhere, and yes, *Machina sapiens* learn faster than we humans in certain areas. But AI systems, for the most part, still do our bidding. The question is, Will things stay this way? In the long run, probably not, but as Lord Keynes said, in the long run, we're all dead. In this book, our interest is the short run: the first half of the twenty-first century, when most of us are still alive. Through this first half of the twenty-first century, AI will be a wonder, new and exciting, with breathtaking breakthroughs—like the best early days of the Internet only better, faster, smarter, and (one hopes) more secure.

If the machines are getting smarter than we are, exponentially smarter, law-of-accelerating-returns smarter—hooking up, sharing data, running statistical analyses, and performing recursive-learning backflips—then AIs are going to come up with things we mere mortals could never have imagined. They'll evolve increasingly toward spontaneity and self-experimentation and away from both human inertia and machine rigidity. They'll avoid the human problem of overthinking and the neural net bugaboo of "overlearning."[3] And every step of the way they'll be obsessed—*obsessed*—with producing better and better results. Because that's just how AIs roll.

If the machines are getting smarter than we are, exponentially smarter, law-of-accelerating-returns smarter, . . . then AIs are going to come up with things we mere mortals could never have imagined.

We've made it thus far without having to stop and formally define our terms. Glossaries and dictionaries make boring reading, but I'm afraid we'll now need to define a few key artificial intelligence terms with some specificity.

The term *artificial intelligence* first emerged at an IBM workshop at Dartmouth College in 1956 and has been the umbrella term for machine learning research and development ever since.[4] In that role, it has not been especially useful. *Artificial intelligence* is overly broad and means different things to different people, and its first name is often a pejorative. There is nothing artificial about AI—at least not in the sense of artificial flowers and artificial smiles. As a new form of intelligence, it's quite real.

The last thing those of us working to improve AI literacy want to do is suggest that *artificial* intelligence is somehow fake or ingenuine. We'd do well to drop the *artificial* part and just speak of AIs and intelligence. The term *AI*, I think, still works well, in the same way *IBM* fits for the company previously known as *International Business Machines*, so *AI* is the term I use most often for this remarkable new intelligence—with *AIs* (plural) being the term of choice for multiple instances swarming around us.

Some in the AI community use the term *synthetic intelligence* to denote the fusion of AI and human intelligence, which is an entire field unto itself. (See the bibliography at the back of this book.) My preferred term for the kind of AI that is capable of synthetic intelligence activities is *Machina sapiens*, which, in my mind at least, connotes a class of whole, integrated entities functioning as intelligent agents.

Swarm intelligence, or *SI*, is another intelligence form, studied in both biology and computer science. SI models the behavior of social insects and animals, ranging from ants to geese. SI is always greater than any one individual's intelligence. Swarm intelligence influenced the architecture and methods of the Internet in a number of ways. In their classic book on SI published in 2001, Kennedy, Eberhart, and Shi predicted the rise of robot swarms that would share tasks and rewards in ways similar to social insects.[5] Their assumption was that the AIs in the robot swarms would individually have low intelligence but that the swarms would be capable of acting and working very intelligently as a whole. Of course, if the robots in a swarm are each highly intelligent and they achieve swarm intelligence collectively, then look out.

Another term I should define here is the one in this book's title: *lightspeed learners*. This is my term for the brightest of the bright, the top AIs who are now, like AlphaZero, achieving remarkable amounts of learning in very short periods of time. These are the AIs we really need to keep our eyes on.

As for AI as an industry, experts divide it into three parts:

1. **Narrow AI (a.k.a., weak AI).** AI today. Idiot savant systems that are highly efficient in narrow pursuits. Narrow AI discovered water on Mars using spectral light signatures and enabled paraplegics to walk.

threat to North America and Europe in the emergence of AIs, its initials are *PRC* (see chapter 4).

For a variety of reasons but most of all because of the uncertainty at the heart of the AI invasion, the more I think about it, the more convinced I become of the need for a new global AI regulatory framework along the lines of the current International Atomic Energy Agency but with substantial differences, as well. When I attempt to figure out, as dispassionately as possible, how this new, not-your-father's generation of AI will play out in the global geopolitical ecosystem, three things become clear:

1. Broadly speaking, AIs are unpredictable and increasingly will have minds of their own. We need a formal network of qualified observers around the world to watch them closely, much as we have done with nuclear, chemical, and biological weapons—and this monitoring must also include commercial uses.
2. It is possible that AI will introduce new existential threats to humanity within the next twenty to thirty years, beyond the threat of humans using them as weapons. It's not likely, in my opinion, but it's possible. AI is software, and getting software to work up to expectations is always harder than it looks. But should we enter an era with general AIs living and learning among us, the potential for serious and dangerous consequences is high enough that we should be taking steps now to ensure that AIs remain safe and friendly.
3. Somehow, some way, we need a Paris Agreement for AI—only stronger. A strategy for getting to such an agreement is presented in the final chapter of this book. Here's the tease: It starts with United States taking aggressive unilateral action to develop a national AI policy, moves quickly to the forging of a major US–China alliance to keep AI safe and friendly, and then links this alliance to the rest of the world in order to create (notionally) a United Nations of AI.

Regardless of whether the United States, China, or any other nation ever reaches such an AI accord, we North Americans and Europeans need to reboot our attitudes about AI—each and every one of us. Experts with heads down building convolutional neural networks, nurses working with smiling robots, supervisors of fleets of self-driving forklifts, search engine AI gurus: We need you all to come up for air occasionally and participate in our national social and political debates on AI. Cybersecurity and system-safety engineers: We need you to jump into policy discussions to make sure that security controls are not overlooked. We need economists, educators,

lawyers, doctors, and storytellers in the AI ethics scrum. Above all, we need AI-literate citizens electing AI-savvy politicians who are ready to address, without political gobbledygook, the fact that we are in the early stages of the biggest technology disruption ever—at least the biggest thing since fire, to steal a line from Google CEO Sundar Pichai.[8]

Making a personal effort to learn more about AI isn't easy. It doesn't help that much of what you know about AI today is probably wrong. I run into people all the time, smart people, who still think of AI as robots. That's like thinking of music as subwoofers.

AI is a global intelligence cloud, an increasingly pervasive grid of connected intelligence. Teilhard de Chardin, the Jesuit mystic anthropologist and discoverer of Peking Man, was first to see it coming. In 1922, he foresaw something he called the noosphere—essentially, a new layer of thought and information around Earth. Here's a portion of what Wikipedia has to say about Teilhard and his noosphere: "For Teilhard, the noosphere is the sphere of thought encircling the earth . . . as much part of nature as . . . the atmosphere, and biosphere." He saw this Thought (he usually capitalized the word) as flowing from man, somehow, into the ether and definitely believed it would evolve its own type of "unified consciousness." Sounds rather like the kind of infosphere a swarm of AGIs might produce.

What's remarkable about Teilhard's vision is how prescient it was, coming precomputer and pre-Internet. Maybe it was all those years spent traversing the steppes of Mongolia, searching for ancient human bones, but by whatever means he got there, Teilhard recognized that Earth was developing a sphere of knowing—to go along with its spheres of carbon organisms, oxygen gas, and others.

Nearly one hundred years later, the mystic Jesuit's "new realm of intelligence" is not only beginning to encircle the globe, but it is also, many believe, headed toward a kind of digital Big Bang that will change everything. Could there really be a technological singularity—a total game-changer—in our future?

CHAPTER THREE

A Leap of Singularities

> "It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers. . . . At some stage therefore, we should have to expect the machines to take control."
>
> —Alan Turing, 1950

To the average onlooker, AI must appear to be one strange technology. Since *2001: A Space Odyssey* fifty years ago, a host of serious films have been made about AI. AIs are popping up everywhere, from hospitals to police cars. AIs are picking the best shots in the US Tennis Open. The world's first AI pet—a doglike robot—is being sold in Japan. In Bangkok, there is an online service for "AI lovers" that seems to be a cross between Match.com and *Her*. And, there is this: AI is the only industry with its own D-Day, something called the *technology singularity*, or the explosive moment when machines become smarter than humans, leading to a dramatic discontinuity in human history.[1]

This AI singularity stuff is spooky. The term comes from astrophysics' *gravitational singularities*, those strange regions in black holes that lie beyond the event horizon, beyond the knowable. These spaces are so tightly packed, so *singular*, that the density of matter becomes infinite. As the eminent American physicist Kip Thorne describes it, a gravitational singularity is the "point where all laws of physics break down."[2] Verner Vinge, the San Diego scientist and science fiction writer who in 1993 popularized *technology singularity* as it is now used in the AI field, said he borrowed the term from

black-hole theory intentionally so that it would convey a proper sense of mystery and dread.[3]

Over the past several decades, the concept of a technology singularity caused by the growth and development of AIs has moved from the realm of science fiction to serious scientific research and study. There has been much focus on the timing of such a singularity, but in my mind, that's less important than its likelihood of coming to pass, on whatever timeline, because, even if such a watershed event is still a century away, there are things we should be doing now to prepare for it, if not prevent it.

The previous paragraph assumes the conventional view that the singularity is a thing—a huge cataclysmic event. But the technology singularity is really just an ongoing thought experiment. Nobody knows what will happen if and when *Machina sapiens* get significantly smarter than us. No human on Earth is close to being genius enough to figure how machines would run the world, if they could and if they wanted to, but that hasn't stopped AI guru Ray Kurzweil from trying.

Kurzweil predicts our planet will reach the technological singularity between 2030 and 2045, and he is not the only one. Whatever the timeline, many highly sapient humans regard the AI singularity as an existential threat. As a prominent Canadian deep-learning professor often reminds his students, the record of less intelligent species retaining control over more intelligent species is not good.

Some AI visionaries see the *technology singularity* as a hostile takeover of human civilization. Other experts, including Kurzweil, believe we humans will glide through this biggest of all tech disruptions and, after perhaps some rough sledding, make it successfully to the other side, where, through some fancy shapeshifting, our minds will live forever, floating on quantum clouds.

The first problem with the *technology singularity* is the unfortunate fact that *singularity* is a word with multiple meanings. It is the state of being singular, peculiar, or unusual; a point where all parallel lines meet; the dense center of a black hole, where light goes to die. In computer science, the singularity is the hypothesis that superintelligent AIs will "trigger runaway technological growth, resulting in unfathomable changes to human civilization."[4] Kevin Kelly, founding editor of *Wired* magazine and author of *Out of Control* (one of the best books ever on digital networks, published 1994), defines the singularity rather chillingly as the point at which "all the change in the last million years will be superseded by the change in the next five minutes."[5]

Yet mention the singularity at an AI industry conference, and all you'll get are eye rolls. My rule of thumb: Among AI experts, concern about a coming technological singularity is inversely proportional to actual, hands-on AI

The singularity can often sound more like a *Star Trek* episode than real science. And try as I might, I just cannot envision Musk's immutable AI dictator as part of our collective future. I wouldn't rule out the possibility of AI-based catastrophes at some point, putting humanity at existential risk, perhaps accidentally, perhaps not. If new "singular" technological events happen, where an AI or two take over some function on their own—if AI monkeys begin running the zoo—then we would cross a major threshold.

Whether crossing such a threshold would stop time and history, as it were, is another question. For instance, if AIs were to get smart enough to take over the running of zoos, will they be friendly or not? To humans, and to animals? There is at least a chance that AIs and robots could develop minds of their own, in both senses of that phrase. Over time, in some situations, their motivations might well clash with human motivations in certain ways. Their motivations could appear evil from a human perspective while being perfectly rational from an AI's point of view. Labeling such motivations as evil is another form of anthropomorphism. Yet, as has been amply demonstrated in the marketplace, evil robots sell books and movie tickets, so I certainly could not leave them out of this book. But killer robots are not anything I expect to be concerned with in my lifetime.

As a thought experiment, let's say that high-tech giants like the GAFA dudes (Google, Amazon, Facebook, Apple) and their allies concluded they had roughly the same stake in denying the technological singularity that oil and gas companies have in denying global climate change. The tech giants would never take the tact of funding pseudoscientists and promoting fake science, but they might promote the idea (in national TV ads) that AI is safe, fun, and helpful or that it can spot tumors, predict elevator failures, and be used to make a golf ball travel farther (all real TV ads today). And—this is the crucial part—they could suggest, as former Google chairman Eric Schmidt and other top execs have done frequently, that worries about AIs running amok are just plain silly. *Don't worry people. We got this.*

The big AI companies have been promoting worry-free AI story lines of late, albeit with less cynicism than the oil and natural gas companies. Jeff Bezos, Eric Schmidt, Tim Cook, and others, in their heart of hearts, I think, really do believe that AI is just another form of software and that it will be a tremendous boon for humanity because, hey, whatever is good for Amazon, Google, and Apple is good for everyone, right?

The highly successful and extremely rich chairmen and CEOs who run Google, Amazon, Facebook, Apple, Microsoft, Intel, NVIDIA, Baidu, Tencent, and Alibaba are all pom-pom-waving, high-kicking, male cheerleaders

for AI. And why not? Not only have their companies made huge AI investments, but these execs understand AI's power. AI is the next big wave of socioeconomic disruption. Time to mount your corporate surfboards, gentlemen, and start paddling. These guys who run high-tech companies feed on tech disruptions the way sea lions feed on salmon, but they are the last people I want making critical and fateful decisions on my behalf about the future of this powerful technology. The impolite question in the industry is, *Who controls AI?* Who sets its goal parameters? Who defines safety and security standards? Who decides how much risk we, as a society, are prepared to take? Amazon and Google or you and me?

In 2017, it was widely reported that a group of AI bots at Facebook developed their own language.[7] These bots were being trained to conduct online negotiations with Facebook's advertising customers. No one at Facebook created this language, and no one understood it. When the bots did communicate in English, they even learned how to lie to Facebook engineers. According to numerous reports, once this minor rebellion was observed, the bot project was shut down immediately. Facebook has neither confirmed nor denied this story officially—and the incident should not be overblown. But this is not the only time AIs have, in effect, hacked themselves and produced unpredictable or untoward results.

In a remarkable and entertaining academic paper published in March 2018, Uber AI researcher Joel Lehman and fifty other scientists and engineers listed twenty-seven anecdotes about unique and unexpected evolutionary behavior on the part of AIs and other cognitive systems.[8] The theory behind the paper is that evolution is as much a force in digital systems as in biological ones, that mutations—surprises—occur that help AIs adapt to their environment and grow. The researchers essentially crowdsourced stories of AI research and development from around the world and then grouped these anecdotes into four sections:

1. **Selection Gone Wild:** The digital evolution of AIs surprises the researchers running experiments. Examples in this category include a robot that learned to do somersaults instead of running and a food-recognition experiment where the AI learned to ignore all sensory data inputs because the researchers always alternated one safe food with one that was poison. Instead of processing data, the AI simply oscillated in order to produce the desired results.
2. **Unintended Debugging:** Digital evolution reveals and exploits previously unknown software or hardware bugs. One example is a game-playing AI that learned to make nonexistent moves in a large

tic-tac-toe board as a way to fry his opponent's memory, causing the opponent to forfeit.

3. **Exceeded Expectations:** Digital evolution produces results that exceed the expectations of experimenters. In this category, a type of "digital organism" invented a step counter on its own in order to stop itself from wandering off its prescribed path.
4. **Convergence with Biology:** Digital evolution "discovers solutions surprisingly convergent with those found in nature, despite vast divergence in medium and conditions. Examples here include various types of mimicry and remarkable results in several genetics experiments.[8]

In this paper, Lehman et al. convincingly make the case, with dozens of well-curated examples, that the element of surprise is a natural part of evolution. And that evolutionary surprise is a part of all complex evolving systems, AIs included.

This means that Jacks will keep springing up out of AI boxes—one-eyed wild-card Jacks producing unanticipated twists and turns; changing AI fitness[9] results, both good and bad; and unexpectedly hacking themselves. According to the authors of this study, these sorts of experimental anomalies usually go unreported, except through informal channels, precisely because they fall outside the testing and training parameters of the project. But when scientists start collecting such anecdotes methodically, a clear pattern emerges—a pattern of cognitive systems behaving in strange and unpredictable ways. Just as with biological evolution, interactions with the environment and successful adaptions to it will drive evolutionary progress in AI. But as evidenced by AlphaZero when learning chess in four hours and becoming the world's best player in sixteen, the learning cycles will come much, much faster than with biological creatures.

Pavlov's AIs

In my experience in the AI industry, the smartest, most intellectually honest scientists always have a healthy respect for the unpredictability of AI. Of course, AI is a bit out of control. This should not come as a big shock. No one completely understands how it works or what it's capable of. So far, AIs are out of control largely in ways that haven't hurt us. The loose, free-market "governance" of AI, if you will, is no doubt at least partially responsible for AI's recent progress because, when there are no laws, you can go as fast as you want. But think out another ten years or so, with massively connected (and heavily funded) global digital ecosystems leveraging several new generations

of semiconductors, cloud systems, Internet sensors, and lightspeed learning algorithms. More and more of our entire societal infrastructure will be riding on these ecosystems for the production and distribution of goods and services of all kinds. What kind of "out-of-control" AI stories will we be hearing then?

All quests to keep AIs friendly will need the cooperation of the AIs themselves. That is to say, the AIs must have the proper goals and motivations. This is especially important in cases where AIs conduct ongoing self-optimization and are engaged in what's known technically as "recursive self-improvement." One AI recursive self-improvement method is called reinforcement learning. It differs from deep learning in that it does not require massive amounts of labeled data. Reinforcement learning uses a method much like the process Ivan Pavlov used to train his famous dogs.

Central to all methods of shaping the behavior of intelligent agents (such as dogs, humans, and robots) is the notion of the "reward function," a payoff of some kind for desired behavior. The smarter and stronger AI becomes, the more AI motivation and reward functions will matter.

Today's AIs have very narrow goals, such as "match user preferences for books" or "detect these faces in surveillance video streams." Even when they are more complicated ("drive car according to all known traffic laws"), the goals are still fairly specific, quite unlike typical human goals, such as "become happier," "prepare for retirement," or "win national election." With respect to this last goal, you can certainly deploy many AIs in service to a national political campaign today, but you cannot flip one AI switch to produce a complete battle plan—at least not yet.

The better AIs become at executing broader goals, the more important it becomes to set the right reward function. And because goals and motivations always change over time as an entity interacts with its environment, various forms of reinforcement learning and recursive learning will play increasing roles, as well. Setting the right goals for AIs requires good logic, judgment, common sense, and transparent biases more than skill in mathematics or software engineering, which is yet another reason it is so important that ordinary humans become prominent in AI industry.

Working in cybersecurity in the mid-2000s, I could never have imagined the theft of three billion identities in one fell swoop ten years later, as happened at Yahoo in 2013. I'm older and less trusting now of the IT industry, so I *can* imagine some fairly horrific AI surprises by 2030, unless we take the kind of action we did not take in cybersecurity—and in climate science—in the 1990s and early 2000s; unless, frankly, AI becomes something a great many of us do together, with science as the wind at our backs and the preservation of our humanity as our north star.

Utilitarians versus Singularitarians

The AI community today is divided into two sects: *AI utilitarians*, whose numbers include most tech company executives, investors, and working AI scientists and engineers. and *AI singularitarians*, the tech cognoscenti who write books, make movies, teach at MIT, and have time to think about the future. Utilitarians believe that AI can be useful to humanity and create economic opportunity safely. They believe that we have decades, if not centuries, to figure out the existential issues related to AGI-level intelligence. Singularitarians believe machines will inevitably become smarter than us—sooner than utilitarians think. The singularity transition point will launch a utopian era, some believe. But most also agree that, if we do not quickly install ethical AI controls, we could enter a dystopian nightmare.

Ray Kurzweil is the world's leading singularitarian. I fully subscribe to Kurzweil's Law of Accelerating Returns but not to his prediction that the technology singularity will arrive by 2030. Ray is a wonderfully visionary thinker, but this is the guy who predicted in 1999 that all computing would dissolve into our eyeglasses and clothing by 2009, that human musicians would be jamming with their cyberequivalents every night, and that we'd all be popping FDA-approved no-fat pills that let us eat everything and anything we want.[10] To be fair, he did predict the fall of the Soviet Union (due to its poor technology) and the rise of a massive global Internet at a time when that network had only two million mostly academic users.

Kurzweil is a brilliant technology promoter and the undisputed singularity guru. As a teenager, he appeared on national television on a popular program called *I've Got a Secret*. His secret was that he had taught a computer to write the music the show's celebrity panel had just heard. He has been a great advocate of computing power ever since. But Kurzweil, who, prior to his recent job at Google, worked mostly in academia and in media, has always had a blind spot: He fails to recognize and appreciate the way real-world inertia slows the train of high-tech innovation.[11]

Over the next decade, AIs will continue to transform markets, disrupt careers, and change lifestyles—even if at slower rates than Kurzweil and the singularitarians predict. They will create winners and losers in business, finance, politics, health care, transportation, and warfare. This is not idle speculation: The data are already trending strongly in this direction. *When* the greatest AI disruptions hit is open to debate, but the fact that a new AI-enabled world is coming is not. Ask Xi Jinping, Vladimir Putin, or Emmanuel Macron.

Some consider AI just the latest shiny new tech object, but it is fundamentally different, a frog's leap into unexplored territory. The inertia of

only reason we stayed with Azure is that Swan had, at my premature suggestion, stopped all of our hardware leases and sold most of the servers. Not the smartest business decision, perhaps, but times were tough. Swan had little choice but to soldier on in Azure.

Then things started settling down. The Azure team instituted a premium customer service that enabled Swan engineers to get customer service from our friends in Redmond directly, the same ones they'd been sharing *kai yaang* with at Typhoon. Service interruptions were still frequent, but redundant failover systems minimized the impact. Holistically, Swan's cloud service started working much better. Through trial and error and various iterations of learning together with the Azure team, performance improved—a lot. So much so that in one stretch beginning in 2013, Swan's 24/7 software service did not have a single minute of unscheduled downtime for more than twenty months.

Suddenly, Swan's TX360 subscription service was running like a Swiss chocolate factory—distributing millions of alerts and thousands of custom dashboards every month to the security pros who protect global multinational corporations. And Azure was on its way to becoming Microsoft's fastest-growing and most valuable business unit, with 93 percent year-over-year growth in 2017.[13]

It is this ability of AIs to learn at lightspeed—and then use this learning autonomously to retune the internal algorithms that do their "thinking"—that makes this AI disruption so unlike any before it.

Technologies that cause major disruptions are always messy in the beginning. New tech sucks until somehow it doesn't. The longer it survives, the more people use it, the better it becomes. This has been true of PCs, printers, databases, phones, and nearly every other digital innovation I can think of. The shakeout period can be lengthy and painful and take years, even decades, yet at the heart of this process is a fairly simple cycle:

1. Build technology product.
2. Market product.
3. Collect customer data about use of the product in marketplace.
4. Use customer feedback about product to design product improvements.
5. Return to step 1.

In the 1990s this cycle generally took several years. In specialty enterprise software markets—the domain of nurse-staffing software, electric-utility regulatory-compliance systems, geological information services for oil drilling, military alerting systems, and the like—a release 2.0 might come five years after the 1.0 launch. It was a long, hard slog to produce a new version—and the relaunches didn't always go well (cf., the history of Microsoft Windows).

Today, largely because of the flexibility and scalability of cloud services, certain non-AI software applications issue new releases *every hour*. These releases for the most part are just small tweaks of the main code base, but the evolutionary cycle is quite fast. Yet however fast, ordinary software does not learn on its own or write itself.

AI is a new class of software. Its release cycles not only come faster, but they also leap ahead prodigiously in performance, one release to the next. MIT professor Max Tegmark has projected a mythical Omega team that issues a new AI release every hour—not a minor tweak but a major new leap. The very thought that software could materially leap forward every hour is, to this veteran of enterprise software wars, astonishing. In my enterprise software career, a traditional software release meant that our product teams collect customer requirements; design new features; write code in a series of short "sprints"; produce alpha, beta, and "release candidate" versions of the product; and then test and fix bugs until we reached the finish line. We called the finish line "code freeze"—and each product's software code did indeed remain frozen for months, even years, after a release. In Tegmark's vision, the Omega code is never frozen. It evolves constantly—continually optimizing, always improving, ever evolving. The better neural networks are beginning to operate this way today, to a degree.

The Omega team is science fiction or, more accurately, science speculation. But Tegmark is right about the ability of AIs to learn and grow quickly. It is this ability of AIs to learn at lightspeed—and then use this learning autonomously to retune their internal algorithms and rewrite their own code—that makes this AI disruption so unlike any technology revolution in history.

If and when we move beyond narrow AIs and reach a point when general and super AIs start cogitating among us, get ready to be seriously intimidated. Imagine trying to win a legal argument with an intelligent machine that has read and digested every legal opinion in the history of American jurisprudence—and that can apply case law and precedent with unerring precision. Imagine playing fantasy football against a competitor who knows the ball-security stats of every current NFL quarterback and running back on wet versus dry fields, and the "fumbles caused" stat line of every defensive