Empire of AI

Dreams and Nightmares in Sam Altman's OpenAI

Karen Hao

Empire of AI

Dreams and Nightmares in Sam Altman's OpenAI

KAREN HAO

PENGUIN PRESS NEW YORK 2025

PENGUIN PRESS
An imprint of Penguin Random House LLC
1745 Broadway, New York, NY 10019
penguinrandomhouse.com

Hardcover ISBN 9780593657508
International edition ISBN 9798217060481
Ebook ISBN 9780593657515

Cover design: Chris Allen
Book design by Daniel Lagin, adapted for ebook by Cora Wigen

The authorized representative in the EU for product safety and compliance is Penguin Random House Ireland, Morrison Chambers, 32 Nassau Street, Dublin D02 YH68, Ireland, https://eu-contact.penguin.ie.

pid_prh_7.1a_151466342_c0_r0

CONTENTS

III

IV

To my family,
past, present, and future.

To the movements
around the world
who refuse dispossession
in the name of abundance.

It is said that to explain is to explain away. This maxim is nowhere so well fulfilled as in the area of computer programming, especially in what is called heuristic programming and artificial intelligence. For in those realms machines are made to behave in wondrous ways, often sufficient to dazzle even the most experienced observer. But once a particular program is unmasked, once its inner workings are explained in language sufficiently plain to induce understanding, its magic crumbles away; it stands revealed as a mere collection of procedures, each quite comprehensible. The observer says to himself "I could have written that." With that thought he moves the program in question from the shelf marked "intelligent," to that reserved for curios, fit to be discussed only with people less enlightened than he.

—JOSEPH WEIZENBAUM, MIT PROFESSOR AND INVENTOR OF THE FIRST CHATBOT, ELIZA, 1966

"Successful people create companies. More successful people create countries. The most successful people create religions."

I heard this from Qi Lu; I'm not sure what the source is. It got me thinking, though--the most successful founders do not set out to create companies. They are on a mission to create something closer to a religion, and at some point it turns out that forming a company is the easiest way to do so.

—SAM ALTMAN, 2013

AUTHOR'S NOTE

This book is based on over 300 interviews with around 260 people and an extensive trove of correspondence and documents. Most of the interviews were conducted for this book. Some were drawn from my last seven years of reporting on OpenAI, the AI industry, and its global impacts for *MIT Technology Review*, *The Wall Street Journal*, and *The Atlantic*. Over 150 of the interviews were with more than 90 current or former OpenAI executives and employees, and a handful of contractors who had access to detailed documentation of parts of OpenAI's model development practices. Another share of the interviews was with some 40 current and former executives and employees at Microsoft, Anthropic, Meta, Google, DeepMind, and Scale, as well as people close to Sam Altman.

Any quoted emails, documents, or Slack messages come from copies or screenshots of those documents and correspondences or are exactly as they appear in lawsuits. In cases where I do not have a copy, I paraphrase the text without quotes. There is one exception, which I mark in the endnotes. All dialogue is reconstructed from people's memories, from contemporaneous notes, or, when marked in the endnotes, pulled from an audio recording or transcript. In most cases, I or my fact-checking team asked those recalling quotes to repeat or confirm them again several months apart to test their stability. Every scene, every number, every name and code name, and every technical detail about OpenAI's models, such as the composition of their training data or the number of chips they were trained on, is corroborated by at least two people, with contemporaneous notes and documentation, or, in a few cases

that I mark in the endnotes, with other media reporting. The same is true for most every other detail about OpenAI in the book. If I named someone, it does not mean I spoke to them directly. When I reference anyone's thoughts or feelings, it is because they described that thought or feeling, either to me, to someone I spoke to, in an email or recording I obtained, or in a public interview.

This book is not a corporate book. While it tells the inside story of OpenAI, that story is meant to be a prism through which to see far beyond this one company. It is a profile of a scientific ambition turned into an aggressive ideological, money-fueled quest; an examination of its multifaceted and expansive footprint; a meditation on power. To that end, in the course of my reporting, I spent significant time embedding with communities on the ground in countries around the world to understand their histories, cultures, lives, and experiences grappling with the visceral impacts of AI. My hope is that their stories shine through in these pages as much as the stories within the walls of one of Silicon Valley's most secretive organizations.

I reached out to all of the key figures and companies that are described in this book to seek interviews and comment. OpenAI and Sam Altman chose not to cooperate.

Prologue

A Run for the Throne

On Friday, November 17, 2023, around noon Pacific time, Sam Altman, CEO of OpenAI, Silicon Valley's golden boy, avatar of the generative AI revolution, logged on to a Google Meet to see four of his five board members staring at him.

From his video square, board member Ilya Sutskever, OpenAI's chief scientist, was brief: Altman was being fired. The announcement would go out momentarily.

Altman was in his room at a luxury hotel in Las Vegas to attend the city's first Formula One race in a generation, a star-studded affair with guests from Rihanna to David Beckham. The trip was a short reprieve in the middle of the punishing travel schedule he had maintained ever since the company released ChatGPT about a year earlier. For a moment, he was too stunned to speak. He looked away as he sought to regain his composure. As the conversation continued, he tried in his characteristic way to smooth things over.

"How can I help?" he asked.

The board told him to support the interim chief executive they had selected, Mira Murati, who had been serving as his chief technology officer. Altman, still confused and wondering whether this was a bad dream, acquiesced.

Minutes later, Sutskever sent another Google Meet link to Greg Brockman, OpenAI's president and a close ally to Altman who had been the only board member missing from the previous meeting.

Sutskever told Brockman he would no longer be on the board but would retain his role at the company.

The public announcement went up soon thereafter. "Mr. Altman's departure follows a deliberative review process by the board, which concluded that he was not consistently candid in his communications with the board, hindering its ability to exercise its responsibilities. The board no longer has confidence in his ability to continue leading OpenAI."

On the face of it, OpenAI had been at the height of its power. Ever since the launch of ChatGPT in November 2022, it had become Silicon Valley's most spectacular success story. ChatGPT was the fastest-growing consumer app in history. The startup's valuation was on the kind of meteoric ascent that made investors salivate and top talent clamor to join the rocket-ship company. Just weeks before, it had been valued at up to $90 billion as part of a tender offer it was in the middle of finalizing that would allow employees to sell their shares to said eager investors. A few days before, it had held a highly anticipated and highly celebrated event to launch its most aggressive slate of products.

Altman was, as far as the public was concerned, the man who had made it all happen. He had spent the spring and summer touring the world, reaching a level of celebrity that was leading the media to compare him to Taylor Swift. He had wowed just about everyone with his unassuming small frame, bold declarations, and apparent sincerity.

Before Vegas, he had once again been globe-trotting, sitting on a panel at the APEC CEO Summit, delivering lines with his usual dazzling effect.

"Why are you devoting your life to this work?" Laurene Powell Jobs, founder and president of the Emerson Collective and Steve

Jobs's widow, had asked him.

"I think this will be the most transformative and beneficial technology humanity has yet invented," he said. "Four times now in the history of OpenAI—the most recent time was just in the last couple of weeks—I have gotten to be in the room, when we sort of push the veil of ignorance back and the frontier of discovery forward, and getting to do that is, like, the professional honor of a lifetime."

Shocked employees learned about Altman's firing just as everyone else did, the link to the public announcement zipping from one phone to the next across the company. It was the chasm between the news and Altman's glowing reputation that startled them the most. The company was by now pushing eight hundred people. These days, employees had fewer opportunities to meet and interact with their CEO in person. But his charming demeanor on global stages was not unlike how he behaved during all-hands meetings, at company functions, and, when he wasn't traveling, around the office.

As the rumor mill kicked into a frenzy and employees doomscrolled X, formerly Twitter, for any shreds of information, someone in the office latched on to what they saw as the most logical explanation and shouted, "Altman's running for president!" It created a momentary release of tension, before people realized this was not the case, and speculation started anew with fresh intensity and dread. Had Altman done something illegal? Maybe it was related to his sister, employees wondered. She had alleged in tweets that had gone viral a month before that her brother had abused her. Maybe it wasn't something illegal but ethically untoward, they speculated, perhaps related to Altman's other investments or his fundraising with Saudi investors for a new AI chip venture.

Sutskever posted an announcement in OpenAI's Slack. In two hours, he would hold a virtual all-hands meeting to answer employee questions. "That was the longest two hours ever," an employee remembers.

Sutskever, Murati, and OpenAI's remaining executives came onto the screen side by side, stiff and unrehearsed, as the all-hands streamed to employees in the office and working from home.

Sutskever looked solemn. He was known among employees as a deep thinker and a mystic, regularly speaking in spiritual terms with a force of sincerity that could be endearing to some and off-putting to others. He was also a goofball and gentlehearted. He wore shirts with animals on them to the office and loved to paint them as well—a cuddly cat, cuddly alpacas, a cuddly fire-breathing dragon—alongside abstract faces and everyday objects. Some of his amateur paintings hung around the office, including a trio of flowers blossoming in the shape of OpenAI's logo, a symbol of what he always urged employees to build: "A plurality of humanity-loving AGIs."

Now, he attempted to project a sense of certainty to anxious employees submitting rapid-fire questions via an online document. But Sutskever was an imperfect messenger; he was not one that excelled at landing messages with his audience.

"Was there a specific incident that led to this?" Murati read aloud first from the list of employee questions.

"Many of the questions in the document will be about the details," Sutskever responded. "What, when, how, who, exactly. I wish I could go into the details. But I can't." Anyone curious should read the press release, he added. "It actually says a lot of stuff. Read it maybe a few times."

The response baffled employees. They had just received cataclysmic news. Surely, as the people most directly affected by the situation, they deserved more specifics than the general public.

Murati read off a few more questions. How did this affect the relationship with Microsoft? Microsoft, OpenAI's biggest backer and exclusive licensee of its technologies, was the sole supplier of its computing infrastructure. Without it, all the startup's work—performing research, training AI models, launching products—would grind to a halt. Murati responded that she didn't expect it to be affected. They had just had a call with Microsoft's chief executive Satya Nadella and chief technology officer Kevin Scott. "They're all very committed to our work," she said.

What about OpenAI's tender offer? Employees with a certain tenure had been given the option to sell what could amount to millions of dollars' worth of their equity. The tender was so soon that many had made plans to buy property, or already had. "The tender—we're, um, we're going to see," Brad Lightcap, the chief operating officer, waffled. "I am in touch with investors leading the tender and some of our largest investors already on the cap table. All have committed their steadfast support to the company."

After several more questions were met with vague responses, another employee tried again to ascertain what Sam had done. Was this related to his role at the company? Or did it involve his personal life? Sutskever once again directed people to the press release. "The answer is actually there," he said.

Murati read on from the document. "Will questions about details be answered at some point or never?"

Sutskever responded: "Keep your expectations low."

As the all-hands continued and Sutskever's answers seemed to grow more and more out of touch, employee unease quickly turned into

anger.

"When a group of people grow through a difficult experience, they often end up being more united and closer to each other," Sutskever said. "This difficult experience will make us even closer as a team and therefore more productive."

"How do you reconcile the desire to grow together through crisis with a frustrating lack of transparency?" an employee wrote in. "Typically truth is a necessary condition for reconciliation."

"I mean, fair enough," Sutskever replied. "The situation isn't perfect."

Murati tried to quell the rising tension. "The mission is so much bigger than any of us," she said.

Lightcap echoed her message: OpenAI's partners, customers, and investors had all stressed that they continued to resonate with the mission. "If anything, we have a greater duty now, I think, to push hard on that mission."

Sutskever again attempted to be reassuring. "We have all the ingredients, all of them: The computer, the research, the breakthroughs are astounding," he said. "When you feel uncertain, when you feel scared, remember those things. Visualize the size of the cluster in your mind's eye. Just imagine all those GPUs working together."

An employee submitted a new question. "Are we worried about the hostile takeover via coercive influence of the existing board members?" Murati read.

"Hostile takeover?" Sutskever repeated, a new edge in his voice. "The OpenAI nonprofit board has acted entirely in accordance to its objective. It is not a hostile takeover. Not at all. I disagree with this question."

That night, several employees gathered at a colleague's house for a party that had been planned before Altman's firing. There were guests from other AI companies as well, including Google DeepMind and Anthropic.

Right before the event, an alert went out to all attendees. "We are adding a second themed room for tonight: 'The no-OpenAI talk room.' See you all!" In the end, few people stayed long in the room. Most people wanted to talk about OpenAI.

Brockman had announced that afternoon that he was quitting in protest. Microsoft's Nadella, who had been furious about being told about Altman's firing only minutes before it happened, had put out a carefully crafted tweet: "We have a long-term agreement with OpenAI with full access to everything we need to deliver on our innovation agenda and an exciting product roadmap; and remain committed to our partnership, and to Mira and the team."

As rumors continued to proliferate, word arrived that three more senior researchers had quit the company: Jakub Pachocki and Szymon Sidor, early employees who had among the longest tenures at OpenAI, and Aleksander Mądry, an MIT professor on leave who had joined recently. Their departures further alarmed some OpenAI employees, a signal of a bleeding out of leadership and talent that could spook investors and halt the tender offer or, worse, ruin the company. At the party, employees grew more and more despondent and agitated. A dissolution of the tender offer would snatch away a significant financial upside to all their hard labor, to say nothing of a dissolution of the company, which would squander so much promise and hard work.

Also that night, the board and the remaining leadership at the company were holding a series of increasingly hostile meetings. After the all-hands, the false projection of unity between Sutskever

and the other leaders had collapsed. Many of the executives who had sat next to Sutskever during the livestream had been nearly as blindsided as the rest of the staff, having learned of Altman's dismissal moments before it was announced. Riled up by Sutskever's poor performance, they had demanded to meet with the rest of the board. Roughly a dozen executives, including Murati and Lightcap, had gathered in a conference room at the office.

Sutskever was dialed in virtually along with the three independent directors: Adam D'Angelo, the cofounder and CEO of the question-and-answer site Quora; Tasha McCauley, an entrepreneur and adjunct senior management scientist at the policy think tank RAND; and Helen Toner, an Australian-born researcher at another think tank, Georgetown University's CSET, or Center for Security and Emerging Technology.

Under an onslaught of questions, the four board members repeatedly evaded making further disclosures, citing their legal responsibilities to protect confidentiality. Several leaders grew visibly enraged. "You're saying that Sam is untrustworthy," Anna Makanju, the vice president of global affairs, who had often accompanied Altman on his global charm offensive, said furiously. "That's just not our experience with him at all."

The gathered leadership pressed the board to resign and hand their seats to three employees, threatening to all quit if the board didn't comply immediately. Jason Kwon, the chief strategy officer, a lawyer who had previously served as OpenAI's general counsel, upped the ante. It was in fact *illegal* for the board not to resign, he said, because if the company fell apart, this would be a breach of the board members' fiduciary duties.

The board members disagreed. They maintained that they had carefully consulted lawyers in making the decision to fire Altman and had acted in accordance with their delineated responsibilities. OpenAI was not like a normal company, its board not like a normal board. It had a unique structure that Altman had designed himself,

giving the board broad authority to act in the best interest not of OpenAI's shareholders but of its mission: to ensure that AGI, or artificial general intelligence, benefits humanity. Altman had long touted the board's ability to fire him as its most important governance mechanism. Toner underscored the point: "If this action destroys the company, it could in fact be consistent with the mission."

The leadership relayed her words back to employees in real time: Toner didn't care if she destroyed the company. Perhaps, many employees began to conclude, that was even her intention. At the thought of losing all of their equity, a person at the party began to cry.

The next day, Saturday, November 18, dozens of people, including OpenAI employees, gathered together at Altman's $27 million mansion to await more news.

The three senior researchers who had quit, Pachocki, Sidor, and Mądry, had met with Altman and Brockman to talk about re-forming the company and continuing their work. To some, word of their discussions increased employee anxiety: A new OpenAI competitor could intensify the instability at the company. To others it offered hope: If Altman indeed founded a new venture, they would leave to go with him.

OpenAI's remaining leadership gave the board a deadline of 5 p.m. Pacific time that day: Reinstate Altman and resign, or risk a mass employee exodus from the company. The board members refused. Through the weekend, they frantically made calls, sometimes in the middle of the night, to anyone on their roster of connections who would pick up. In the face of mounting ire from employees and investors over Altman's firing, Murati was no longer willing to serve as interim CEO. They needed to replace her with

someone who could help restore stability, or find new board members who could hold their own against Altman if he actually came back.

That night, after the deadline came and went, Jason Kwon sent a memo to employees. "We are still working towards a resolution and we remain optimistic," he wrote. "By resolution, we mean bringing back Sam, Greg, Jakub, Szymon, Aleksander."

Altman tweeted in his signature lowercase style. "i love the openai team so much."

Dozens of other employees began retweeting it with a heart emoji.

On Sunday, Altman and Brockman arrived back at the office to negotiate their return. Over the course of the day, more and more employees joined them to wait in suspense. By then, most employees, leadership, and the board had barely slept in more than thirty-six hours; everything was beginning to blur together. Altman tweeted a selfie, lips pursed, brows furrowed, displaying a guest badge in his hand. "first and last time i ever wear one of these," he added as the caption. Leadership set another 5 p.m. deadline for the board to reinstate Altman and to resign.

The pressure was now piling on from all directions. Microsoft, OpenAI's other investors, and heavyweights across Silicon Valley were publicly siding with Altman. A source relayed the playbook to the media: Not only would employees leave en masse if the decision were not reversed, but Microsoft would withhold access to its computing infrastructure, and investors would file lawsuits. The combination would make an OpenAI without Altman untenable.

Still, the board continued to resist. Nearing 9 p.m., once again well past the latest deadline, Sutskever posted a long message on Slack on behalf of the board. Altman was not returning; Emmett

Shear, the former CEO of Twitch, was now the new interim head of OpenAI. He and Shear would arrive at the office in five minutes to give a speech about the company's new vision.

"The board firmly stands by its decision as the only path to advance and defend the mission of OpenAI," he wrote. "Put simply, Sam's behavior and lack of transparency in his interactions with the board undermined the board's ability to effectively supervise the company in the manner it was mandated to do."

The Slack instantly lit up with dozens of angry replies from employees.

"You and what fucking army"

"you're delusional"

"Emmett will be the CEO of nothing"

Roughly two hundred employees paraded out of the office to boycott the talk. Murati rushed the executives out the building. By the time Shear arrived with Sutskever, only a dozen or so people were in the audience.

Anna Brockman, Greg's wife, approached Sutskever, who four years earlier had officiated the couple's civil ceremony. Through tears, she flung her arms around him and pleaded with him to reconsider his position.

Many of the employees who had left the office gathered at a few colleagues' houses to weather the night; hundreds joined a Signal group for updates. Late that evening, Nadella announced that he was hiring Altman and Brockman to lead a new AI division. Word spread rapidly: Anyone who wanted to join Altman would have a guaranteed job at Microsoft.

The news flipped the mood from fear to defiance. With the perception of a backup option in hand, employees had new leverage to speak out against the board and Shear. At one employee's house,

overflowing with well over a hundred OpenAI colleagues, executives and senior researchers wrote an open letter to amp up the pressure, reiterating the leadership team's threats with greater force: Without Altman's reinstatement and the board's resignation, they could all quit immediately and join Microsoft.

The group worked to circulate the letter as far and wide as possible, posting it on various private channels and phoning employees who were not present to sign it. As it reached a critical mass of signatures, many more employees rushed to join in, under pressure to avoid raising questions about their absence. Within twenty-four hours, the letter had reached more than 700 signatories of the roughly 770 employees. Dozens of employees sent identical emails to the board in rapid succession. In droves, they took to X to post the same message: "OpenAI is nothing without its people."

Then, in the middle of the night, employees saw Sutskever's name appear on the open letter.

Sutskever soon addressed it publicly. "I deeply regret my participation in the board's actions," he tweeted in the early hours of Monday morning. "I never intended to harm OpenAI. I love everything we've built together and I will do everything I can to reunite the company."

On Tuesday, November 21, leadership dialed the board from Altman's house. Five days in, everyone was caving from the lack of sleep and exhaustion. Thanksgiving was around the corner, and desperation on both sides had finally opened a pathway toward a resolution.

Through the day, everyone began to consider different configurations in earnest.

Altman and Brockman, originally adamant about returning to OpenAI with board seats, finally acquiesced to no longer having

them. The board, seeing no path for preserving the company without Altman, finally acquiesced to his return.

Late that night, they agreed on the three independent board seats. D'Angelo would stay; Toner and McCauley would step down; Bret Taylor, a former co-CEO of Salesforce and former CTO of Facebook, and Larry Summers, a former treasury secretary and former president of Harvard, would fill their vacancies. As part of the deal, the new board would eventually add more members. Altman would submit to an investigation.

What was most important for the company now, they also agreed, was to project unity, stability, and reconciliation. Two days later, Altman would tweet a staged message: "just spent a really nice few hours with @adamdangelo. happy thanksgiving from our families to yours 🦃" Ten days later, Brockman would tweet a photo: him and Sutskever, arms around each other, smiling widely. In the office, the company's artist in residence would hook up OpenAI's image generator DALL-E to a color printer to create tiny kaleidoscopic heart-shaped stickers. Next to the printer would be a giant pink heart emblazoned with the line "OpenAI is nothing without its people."

In December, Altman would describe the experience to Trevor Noah in a podcast as the second worst moment of his life, surpassed only by his dad's death. The following month, in January, the tender would close, valuing OpenAI at $86 billion.

But that was all to come. Tuesday night, November 21, was just about celebrating. With the announcement of Altman's return and the new agreement, employees came flooding back to the office to hug, to cry, to blast music. At some point, someone turned on a smoke machine. It set off the fire alarm. Everyone kept partying.

Brockman snapped a group selfie with the crowd, a picture bursting with the ecstatic, slaphappy delirium of surviving a crisis. He tweeted it with a caption: "we are so back."

The news of Altman's ouster broke as I was in the middle of an interview for this book. I had silenced my phone, blissfully unaware of the chaotic week about to unfold. Twenty minutes later, I tapped my screen to check the time and saw a slew of missed notifications. So began a hazy, adrenaline-fueled series of days as I raced to understand what was going on.

In the weeks that followed, friends, family, and media would ask me dozens of times: What did all this mean, if anything? Was the back-and-forth just an entertaining distraction? Or would it have consequences for the rest of us? I had by then been following OpenAI for five years. In 2019, I was the first journalist to gain extensive access to the company and to write its first profile. To me, these events were not just some frivolous Silicon Valley power moves. The drama highlighted one of the most urgent questions of our generation: How do we govern artificial intelligence?

AI is one of the most consequential technologies of this era. In a little over a decade, it has reformed the backbone of the internet, becoming a ubiquitous mediator of digital activities. In even less time, it is now on track to rewire a great many other critical functions in society, from health care to education, from law to finance, from journalism to government. The future of AI—the shape that this technology takes—is inextricably tied to our future. The question of how to govern AI, then, is really a question about how to ensure we make our future better, not worse.

From the beginning, OpenAI had presented itself as a bold experiment in answering this question. It was founded by a group including Elon Musk and Sam Altman, with other billionaire backers like Peter Thiel, to be more than just a research lab or a company. The founders asserted a radical commitment to develop so-called artificial general intelligence, what they described as the most

powerful form of AI anyone had ever seen, not for the financial gains of shareholders but for the benefit of humanity. To that end, Musk and Altman had set it up as a nonprofit and pledged $1 billion for its operation. It would not work on commercial products; instead it would be dedicated fully to research, driven by only the purest intentions of ushering in a form of AGI that would unlock global utopia, and not its opposite. Musk and Altman also pledged to share as much of its research as possible along the way and to collaborate widely with other institutions. If the goal was to do good by the world, openness—hence *Open*AI—and democratic participation in the technology's development were key. A few years later, leadership went even further, making a promise to self-sacrifice if necessary. "We are concerned about late-stage AGI development becoming a competitive race without time for adequate safety precautions," they wrote. If another attempt to create beneficial AGI surpassed OpenAI's progress, "we commit to stop competing with and start assisting this project."

But by the time I began to profile OpenAI, its commitment to these ideals were fast eroding. Merely a year and a half in, OpenAI's executives realized that the path they wanted to take in AI development would demand extraordinary amounts of money. Musk and Altman, who had until then both taken more hands-off approaches as cochairmen, each tried to install himself as CEO. Altman won out. Musk left the organization in early 2018 and took his money with him. In hindsight, the rift was the first major sign that OpenAI was not in fact an altruistic project but rather one of ego.

The loss of its primary backer pushed OpenAI into financial uncertainty. To plug the hole, Altman reformulated OpenAI's legal structure. Nested within the nonprofit, he created a for-profit arm, OpenAI LP, to raise capital, commercialize products, and provide returns to investors much like any other company. Four months later,

in July 2019, OpenAI announced a new $1 billion funder: software giant and cloud services provider Microsoft.

I arrived at OpenAI's offices for the first time shortly thereafter, in August 2019. After three days embedded among employees and dozens of interviews, I could see that the experiment in idealistic governance was unraveling. OpenAI had grown competitive, secretive, and insular, even fearful of the outside world under the intoxicating power of controlling such a paramount technology. Gone were notions of transparency and democracy, of self-sacrifice and collaboration. OpenAI executives had a singular obsession: to be the first to reach artificial general intelligence, to make it in their own image.

Over the next four years, OpenAI became everything that it said it would not be. It turned into a nonprofit in name only, aggressively commercializing products like ChatGPT and seeking unheard-of valuations. It grew even more secretive, not only cutting off access to its own research but shifting norms across the industry to bar a significant share of AI development from public scrutiny. It triggered the very race to the bottom that it had warned about, massively accelerating the technology's commercialization and deployment without shoring up its harmful flaws or the dangerous ways that it could amplify and exploit the fault lines in our society. Along the way, clashes between leaders and employees grew ever more fierce, as different groups inside the company sought to seize control and re-form OpenAI around their vision.

The ouster and reinstatement of Altman in November 2023 was final proof that the governance experiment had failed. Not simply because OpenAI's nonprofit board buckled under moneyed interests, dissolving the last remnant of the organization's altruistic facade. It illustrated in the clearest terms just how much a power struggle among a tiny handful of Silicon Valley elites is shaping the future of AI. Even if events had gone a different way and the board had succeeded in replacing Altman, nothing would have changed about

the fact that such a consequential decision was made behind closed doors. Beyond a small group of ultrarich techno-optimists, their fiercest ideological rivals, and a multibillion-dollar tech giant, even OpenAI's own employees found themselves largely in the dark about which way their fates would fall.

I began reporting on artificial intelligence long before OpenAI and ChatGPT became synonymous with the technology. I watched it evolve through the messy process of science and innovation as researchers trialed new ideas, presented their best successes at packed conferences, and brought them to bear on commercial products at the world's biggest companies, including Google and Facebook, Alibaba and Baidu. I read hundreds of research papers and interviewed scientists, engineers, and executives to understand their worldviews and their decisions—and how those left fingerprints on the technology's design and application.[*] As AI's footprint sprawled out globally, I tracked the subtle and dramatic ways it changed lives and communities. I traveled to five continents to hear from people about these experiences. In Colombia and Kenya, I met people who in the face of economic crisis turned to annotating data for the AI industry, only to find themselves working under conditions that resembled indentured servitude. In Arizona and Chile, I met with local politicians and activists worried about the growing shadow metropolis of data centers guzzling their homes' precious water resources.

Through my reporting, I've come to understand two things: Artificial intelligence is a technology that takes many forms. It is in fact a multitude of technologies that shape-shift and evolve, not merely based on technical merit but with the ideological drives of the people who create them and the winds of hype and commercialization. While ChatGPT and other so-called large language models or generative AI applications have now taken the limelight, they are but one manifestation of AI, a manifestation that embodies a particular and remarkably narrow view about the way

the world is and the way it should be. Nothing about this form of AI coming to the fore or even existing at all was inevitable; it was the culmination of thousands of subjective choices, made by the people who had the power to be in the decision-making room. In the same way, future generations of AI technologies are not predetermined. But the question of governance returns: Who will get to shape them?

The other thing I've learned: This current manifestation of AI, and the trajectory of its development, is headed in an alarming direction. On the surface, generative AI is thrilling: a creative aid for instantly brainstorming ideas and generating writing; a companion to chat with late into the night to ward off loneliness; a tool that could perhaps one day be so effective at boosting productivity that it will increase top-line economic activity. But in the same way we once thought Facebook was merely a place for posting vacation pictures and connecting with long-lost elementary school friends, or for sparking positive and transformative social movements, there is more to the sleek, entrancing exterior than meets the eye. Under the hood, generative AI models are monstrosities, built from consuming previously unfathomable amounts of data, labor, computing power, and natural resources. GPT-4, the successor to the first ChatGPT, is, by one measure, reportedly over fifteen thousand times larger than its first generation, GPT-1, released five years earlier. The exploding human and material costs are settling onto wide swaths of society, especially the most vulnerable, people I met around the world, whether workers and rural residents in the Global North or impoverished communities in the Global South, all suffering new degrees of precarity. Rarely have they seen any "trickle-down" gains of this so-called technological revolution; the benefits of generative AI mostly accrue upward.

Over the years, I've found only one metaphor that encapsulates the nature of what these AI power players are: empires. During the long era of European colonialism, empires seized and extracted

resources that were not their own and exploited the labor of the people they subjugated to mine, cultivate, and refine those resources for the empires' enrichment. They projected racist, dehumanizing ideas of their own superiority and modernity to justify —and even entice the conquered into accepting—the invasion of sovereignty, the theft, and the subjugation. They justified their quest for power by the need to compete with other empires: In an arms race, all bets are off. All this ultimately served to entrench each empire's power and to drive its expansion and progress. In the simplest terms, empires amassed extraordinary riches across space and time, through imposing a colonial world order, at great expense to everyone else.

The empires of AI are not engaged in the same overt violence and brutality that marked this history. But they, too, seize and extract precious resources to feed their vision of artificial intelligence: the work of artists and writers; the data of countless individuals posting about their experiences and observations online; the land, energy, and water required to house and run massive data centers and supercomputers. So too do the new empires exploit the labor of people globally to clean, tabulate, and prepare that data for spinning into lucrative AI technologies. They project tantalizing ideas of modernity and posture aggressively about the need to defeat other empires to provide cover for, and to fuel, invasions of privacy, theft, and the cataclysmic automation of large swaths of meaningful economic opportunities.

OpenAI is now leading our acceleration toward this modern-day colonial world order. In the pursuit of an amorphous vision of progress, its aggressive push on the limits of scale have set the rules for a new era of AI development. Now every tech giant is racing to out-scale one another, spending sums so astronomical that even they have scrambled to redistribute and consolidate their resources. Around the time Microsoft invested $10 billion in OpenAI, it laid off ten thousand workers to cut costs. After Google watched OpenAI

outpace it, it centralized its AI labs into Google DeepMind. As Baidu raced to develop its ChatGPT equivalent, employees working to advance AI technologies for drug discovery had to suspend their research and cede their computer chips to develop the chatbot instead. The current AI paradigm is also choking off alternative paths to AI development. The number of independent researchers not affiliated with or receiving funding from the tech industry has rapidly dwindled, diminishing the diversity of ideas in the field not tied to short-term commercial benefit. Companies themselves, which once invested in sprawling exploratory research, can no longer afford to do so under the weight of the generative AI development bill. Younger generations of scientists are falling in line with the new status quo to make themselves more employable. What was once unprecedented has become the norm.

Today, the empires have never been richer. As I finished writing this book in January 2025, OpenAI topped a $157 billion valuation. Anthropic, a competitor, was nearing a deal that would value it at $60 billion. After striking its partnership with OpenAI, Microsoft tripled its market capitalization to over $3 trillion. Since ChatGPT, the six largest tech giants together have seen their market caps increase $8 trillion. At the same time, more and more doubts have risen about the true economic value of generative AI. In June 2024, a Goldman Sachs report noted spending on the technology's development was projected to hit $1 trillion in a few years with so far "little to show for it." The following month, a survey from The Upwork Research Institute of 2,500 workers globally found that while 96 percent of C-suite leaders expected generative AI to boost productivity, 77 percent of the employees actually using the tools reported them instead adding to their workload; this was in part due to the amount of time spent reviewing AI-generated content, in part due to growing demands from superiors to do more work. In a November *Bloomberg* article reviewing the financial tally of generative AI impacts, staff writers Parmy Olson and Carolyn

Silverman summarized it succinctly—the data "raises an uncomfortable prospect: that this supposedly revolutionary technology might never deliver on its promise of broad economic transformation, but instead just concentrate more wealth at the top."

Meanwhile, the rest of the world is beginning to collapse under the weight of the exploding human and material costs of this new era. Workers in Kenya earned starvation wages to filter out violence and hate speech from OpenAI's technologies, including ChatGPT. Artists are being replaced by the very AI models that were built from their work without their consent or compensation. The journalism industry is atrophying as generative AI technologies spawn heightened volumes of misinformation. Before our eyes, we're seeing an ancient story repeat itself—and this is only the beginning.

OpenAI is not slowing down. It is continuing to chase even greater scales with unparalleled resources, and the rest of the industry is following. To quell the rising concerns about generative AI's present-day performance, Altman has trumpeted the future benefits of AGI ever louder. In a September 2024 blog post, he declared that the "Intelligence Age," characterized by "massive prosperity," would soon be upon us, with superintelligence perhaps arriving as soon as in "a few thousand days." "I believe the future is going to be so bright that no one can do it justice by trying to write about it now," he wrote. "Although it will happen incrementally, astounding triumphs—fixing the climate, establishing a space colony, and the discovery of all of physics—will eventually become commonplace." At this point, AGI is largely rhetorical—a fantastical, all-purpose excuse for OpenAI to continue pushing for ever more wealth and power. Few others have the comparable capital to invest in alternative options. OpenAI and its small handful of competitors will have an oligopoly on the technology they're selling us as the key to the future; anyone—whether company or government—who wants a piece of that vision will have to rely on the empires to provide it.

There is a different way forward. Artificial intelligence doesn't have to be what it is today. We don't need to accept the logic of unprecedented scale and consumption to achieve advancement and progress. So much of what our society actually needs—better health care and education, clean air and clean water, a faster transition away from fossil fuels—can be assisted and advanced with, and sometimes even necessitates, significantly smaller AI models and a diversity of other approaches. AI alone won't be enough, either: We'll also need more social cohesion and global cooperation, some of the very things being challenged by the existing vision of AI development.

But the empires of AI won't give up their power easily. The rest of us will need to wrest back control of this technology's future. And we're at a pivotal moment when that's still possible. Just as empires of old eventually fell to more inclusive forms of governance, we, too, can shape the future of AI together. Policymakers can implement strong data privacy and transparency rules and update intellectual property protections to return people's agency over their data and work. Human rights organizations can advance international labor norms and laws to give data labelers guaranteed wage minimums and humane working conditions as well as to shore up labor rights and guarantee access to dignified economic opportunities across all sectors and industries. Funding agencies can foster renewed diversity in AI research to develop fundamentally new manifestations of what this technology could be. Finally, we can all resist the narratives that OpenAI and the AI industry have told us to hide the mounting social and environmental costs of this technology behind an elusive vision of progress.

Skip Notes

* A note on AI research paper conventions and peer review: In the AI field, researchers often post their papers directly online to a free and open repository called arXiv (pronounced "archive") and either go through a peer-review process with a conference

or publication many months or years later, or do not bother to get peer reviewed at all. This practice has become so normalized that many people in the field cite papers based on impact rather than whether they have passed peer review. In this book, I will do the same. The endnotes denote which papers did not get peer-reviewed as preprints.