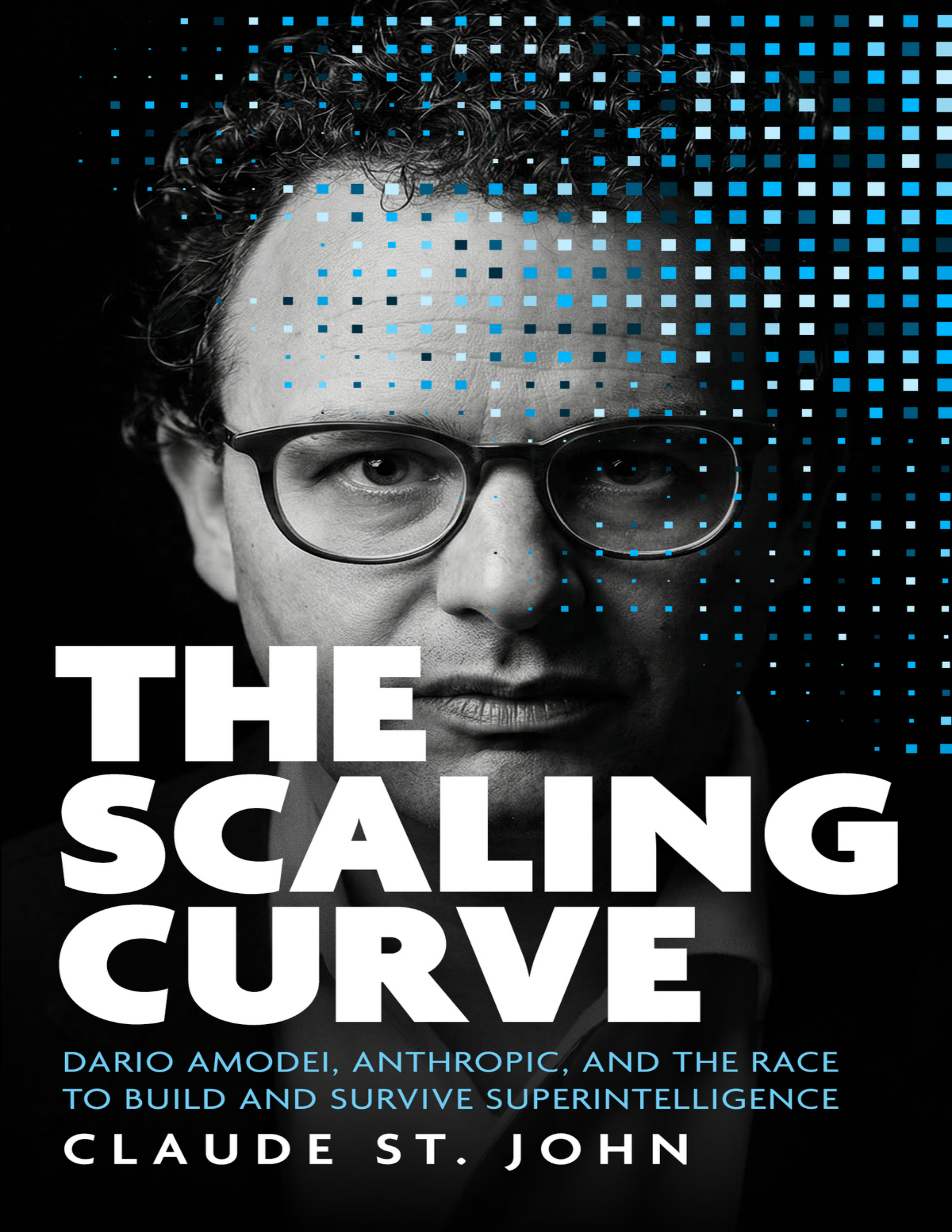

# THE SCALING CURVE

DARIO AMODEI, ANTHROPIC, AND THE RACE TO BUILD AND SURVIVE SUPERINTELLIGENCE

CLAUDE ST. JOHN

# THE SCALING CURVE

Copyright © 2026 by Claude St. John

All rights reserved.

No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without written permission from the author, except for the use of brief quotations in a book review.

ISBN 979-8-248-96654-7 (paperback)

First Edition

# THE SCALING CURVE

—

*Dario Amodei, Anthropic, and the Race to Build and Survive Superintelligence*

Claude St. John

# C O N T E N T S

# CHAPTER ONE

## The Objective Answer

*Childhood, Family, and the Pull of Mathematics*

One kid says the show is great. Another kid says it's terrible. There is no way to settle the argument. But then there is math.

For Dario Amodei, growing up in the Mission District of San Francisco in the late 1980s and early 1990s, this was the discovery that organized his childhood. The Mission in those years was a different place than the neighborhood it would become. The Victorian row houses had not yet been renovated into three-million-dollar homes. The storefronts along Valencia Street still belonged to taquerías, laundromats, and bodegas rather than artisanal coffee shops and co-working spaces. The tech boom that would remake the city, and eventually consume it, was years away. "The city when I first grew up here had not really gentrified that much," Dario would recall. "The tech boom hadn't happened yet. It happened as I was going through high school."

In a rare public discussion of his early years, he tried to explain the appeal of numbers. "I was always really interested in math," he said. "It felt like it had a sense of objectivity. One kid could say, 'Oh, the show is great,' and the other kid could say, 'Oh, it's terrible.' But when you're doing math, you're like—oh man, there's an objective answer to this."

It was a modest observation, the kind of thing any mathematically inclined child might notice. But the instinct it revealed, that the hardest problems have answers waiting to be found, that what matters is finding the right tool to extract them, would shape everything Dario Amodei went on to do. It would lead him to physics, to neuroscience, to the most consequential discovery in modern artificial intelligence. And it would eventually lead him to a question for which, he would come to believe, the objective answer might determine the future of the human species.

• • •

Dario was born in 1983, the first child of Riccardo Amodei and Elena Engel. They were, on paper, an improbable couple. Riccardo was an Italian leather craftsman who had immigrated to the United States from the island of Elba—the same rocky outcrop off the Tuscan coast where Napoleon had spent his first exile. For centuries a modest population had sustained itself there on iron mining, fishing, and the kind of artisanal trades that Riccardo practiced. He had come to San Francisco sometime in the 1970s, part of a long tradition of Italian craftsmen drawn to a city with deep roots in European immigrant communities. Elena was Jewish-American, born and raised in Chicago. She worked as a project manager overseeing renovation and construction for libraries in Berkeley and San Francisco—a job that required the kind of organizational rigor and institutional patience that would, in time, prove hereditary. They settled in the Mission, which suited them: a neighborhood of immigrant families and working people, not yet touched by the wealth that would soon overwhelm it.

The household that Riccardo and Elena built was, in some ways, a study in complementary temperaments. He was the craftsman, working with physical materials, shaping leather into objects of beauty and utility. She was the manager, working with institutions, navigating the bureaucracies of public construction. Together they created an environment that was defined less by any particular intellectual program than by a pervasive moral seriousness. They were not academics. They did not fill the house with books about artificial intelligence or push their children toward careers in technology. What they gave their children was something more fundamental.

Four years after Dario, in 1987, his sister Daniela was born. The age gap was large enough that they were never in the same phase of childhood at the same time. When Dario was twelve and obsessed with physics, Daniela was eight and taking up classical flute. But it was small enough that they grew up genuinely close, in the way that siblings who share not just a household but a temperament can be. Where Dario was drawn to science and mathematics, Daniela studied English literature at the University of California, Santa Cruz. Where he was analytical, she was organizational. But they shared something more fundamental than intellectual interests: an ambition, almost grandiose in its scope, to do something that mattered. They attended Lowell High School, a prestigious public magnet school on the western edge of the city that drew the most academically motivated students from across San Francisco's diverse neighborhoods, though never at the same time; Dario graduated before Daniela arrived. By their own account, they decided early that they wanted to work on something important together. "We always wanted to save the world together," Dario has said, with a self-aware half-smile that suggests he knows how the line sounds. "It's actually kind of amusing that we're working on something together that, at least potentially, could have very, very wide scope."

In the telling of their story, which both siblings have shared publicly on several occasions, the ambition came first and the vehicle came later. Many siblings dream of working together; few execute on it. Dario and Daniela would not find their shared project until decades later. Their careers took them through different worlds: his through the laboratories of Princeton and the AI research groups of Baidu and Google. Hers through a congressional campaign in Pennsylvania, a stint managing communications for a member of Congress, and a career at the payments startup Stripe, where she was an early employee. That their paths converged at one of the defining companies of the twenty-first century was not, from the vantage point of their childhood, predictable in the slightest. What was predictable was that if an opportunity arose to work together on something of sufficient scope, they would take it.

The Amodei household was a loving one. Dario has described it as organized around a set of values that he absorbed so thoroughly he would carry them, largely unexamined, into middle age. “They gave me a sense of right and wrong and what was important in the world,” he has said of his parents. “Imbuing a strong sense of responsibility is maybe the thing that I remember most. They were always people who felt that sense of responsibility and wanted to make the world better.” Neither parent would have guessed that their children would one day stand at the center of the most consequential technological revolution since the splitting of the atom.

But the household did prize a certain kind of seriousness. Dario was, even as a child, drawn to questions with definite answers. He liked physics for the same reason he liked math: there were right answers, and you could find them. The subjective world, of opinion, of taste, of the kind of social maneuvering that occupies most teenagers, held little appeal. He was interested “almost entirely in math and physics.” When the first dot-com boom erupted around him during high school, turning San Francisco into the epicenter of a gold rush that made millionaires of twenty-three-year-olds with half-baked business plans, he barely noticed. “Writing some website actually had no interest to me whatsoever,” he would recall. “Founding a company—those weren’t things that I was interested in at all. I was interested in discovering fundamental scientific truth.”

Dario Amodei did not have the typical origin story of a Silicon Valley CEO. No garage, no lemonade stand, no childhood business selling baseball cards. Dario Amodei wanted to be a scientist. That was the whole of it. And in a culture that celebrated the dropout founder and the teenage disruptor, it was an origin story that turned out to be not a weakness but a quietly radical advantage. The world was about to be changed by a technology whose trajectory could be predicted only by people who understood exponential curves, who could read a loss function the way a cardiologist reads an EKG, and who had spent enough time in laboratories to know the difference between a genuine discovery and a statistical fluke. Dario was all of those things. He just didn't know it yet.

• • •

He was good enough at physics to make the U.S. Physics Olympiad Team in 2000. This placed him in the top tier of high school science students in the country. When it came time for college, he went to Caltech—the California Institute of Technology, in Pasadena—a school whose culture was as distinctive as its location. Caltech was tiny, intense, and almost comically specialized. With fewer than a thousand undergraduates, it had more Nobel Prize winners per capita than perhaps any institution on earth. The students were not well-rounded in the way that Harvard or Yale might have preferred; they were, overwhelmingly, people who had been the best math or science student in their high school and had no particular interest in being anything else. Caltech was a greenhouse for a certain type of mind, and Dario thrived in it. He worked with Professor Tom Tombrello, one of the senior faculty members, as part of Physics 11, a small and somewhat legendary course that paired undergraduates with working researchers on real problems. Not textbook exercises but open-ended questions that had no answer key.

Caltech also revealed a side of Dario that would recur throughout his career: a moral combativeness that could catch people off guard. When the United States invaded Iraq in 2003, Dario vocally confronted his fellow students about what he saw as their passivity. The protest was not especially successful—Caltech undergraduates were not, as a demographic, highly susceptible to political agitation—but it demonstrated something about the way Dario processed the world. Science and morality were not, for him, separate domains. Finding the right answer was a responsibility, not merely an intellectual exercise. And when the people around him were getting the answer wrong, he felt compelled to say so.

During his undergraduate years, Dario transferred from Caltech to Stanford University, where he completed his physics degree. Stanford put him in a different ecosystem: larger, less monastic, more connected to the world of applied science and, inevitably, to the technology industry that pulsed just beyond the campus gates. Caltech was a monastery for pure science: austere, intense, largely indifferent to the commercial world. Stanford was a cathedral at the center of a city, its research programs interwoven with the venture capital firms on Sand Hill Road and the corporate campuses that lined Highway 101. The culture was different, too: at Caltech, the implicit message was that the highest calling was understanding the universe. At Stanford, the implicit message was that understanding the universe was the first step toward changing it. It was at Stanford, or perhaps in the transition between the two schools, that Dario encountered the book that changed the trajectory of his life.

• • •

Ray Kurzweil was, depending on whom you asked, either a visionary futurist or a particularly articulate crank. A prolific inventor who had built pioneering text-to-speech systems and music synthesizers, he had in 1999 published *The Age of Spiritual Machines* and would follow it in 2005 with *The Singularity Is Near*, a book that argued—with graphs and charts and an engineer's confidence in extrapolation—that the exponential growth of computing power would inevitably produce artificial intelligence surpassing human cognition. Moore's Law—the observation that the number of transistors on a chip doubled roughly every eighteen months—was not, in Kurzweil's telling, just a fact about semiconductors. He saw a fact about the universe. Computing power was on an exponential curve. That curve would not stop. And at some point, it would cross the threshold of human-level intelligence.

Much of the scientific establishment regarded Kurzweil's predictions with skepticism bordering on derision. Artificial intelligence, by the early 2000s, had been through two so-called "AI winters"—periods in which funding dried up and the field's grand promises went conspicuously unfulfilled. The first winter, in the 1970s, had followed the failure of early machine translation and symbolic reasoning programs to live up to the sweeping claims made by their creators. The second, in the late 1980s and early 1990s, had come after expert systems—rule-based programs that attempted to encode human knowledge explicitly—proved brittle, expensive, and unable to generalize. By the time Dario was reading Kurzweil as an undergraduate, the very phrase "artificial intelligence" carried a faint stigma in serious computer science departments. Researchers who worked on neural networks often avoided the term altogether, preferring "machine learning" or "statistical learning theory." The idea that raw computing power would somehow, on its own, produce intelligence struck many researchers as naive. Intelligence was not just computation; it was architecture, algorithms, representation, meaning. You could make a calculator infinitely fast and it would never write a poem.

Dario found the argument convincing anyway. Not all of it. He has said that he thought Kurzweil was "a bit crazy about a lot of things," but the core premise seized him. An exponential acceleration of compute was underway. Real and measurable. And it was going to provide enough raw computational power to produce very powerful AI. The mechanism was unclear: "We had no idea how. We had no idea it was going to be neural nets." But the trajectory felt right. The math, as Dario might have put it, had an objective answer.

He had been planning to pursue theoretical physics in graduate school. The default path for a student of his caliber and temperament: a PhD, a postdoc, a career spent pushing the boundary of human understanding of the physical universe. But the Kurzweil revelation redirected him. If artificial intelligence was coming, not as science fiction but as an extrapolation of existing trends, then the most important question in the world was how intelligence worked. And the best available model of intelligence was not a computer. It was a brain.

"I was about to start grad school for theoretical physics," Dario has recalled, "and decided, as soon as I got there, that I wanted to do biophysics and computational neuroscience. Because, you know, if there was going to be AI—it didn't feel like AI was working yet—and so I wanted to study the closest thing to [it] that there was, which was our brains. The natural intelligence. Therefore the closest thing to an artificial intelligence that exists."

He enrolled at Princeton University, one of the strongest programs in the country for the kind of work he wanted to do: applying the mathematical frameworks of physics to the messy, complicated, poorly understood circuitry of the biological brain.

• • •

Dario was awarded a Hertz Fellowship from the Fannie and John Hertz Foundation, one of the most prestigious and competitive fellowships in American science, given to roughly fifteen students per year across all fields of applied physical science. He also received an NDSEG fellowship from the Department of Defense. These were not awards given to merely competent students; they were bets placed on people expected to change the trajectory of their fields. Dario joined the research group of Professor Michael Berry, whose work focused on neural computation: the effort to understand how biological neural circuits process information using the mathematical tools of physics.

What followed were years of the kind of painstaking, detail-oriented work that characterizes experimental biophysics and that rarely appears in accounts of scientific breakthroughs. Dario built devices for intracellular and extracellular recording—instruments designed to measure the electrical activity of individual neurons, both from inside the cell and from the surrounding tissue. He developed statistical mechanics models of neural circuits, borrowing mathematical frameworks from thermodynamics and applying them to systems of interconnected neurons. The work was technical, elegant in its methodology and representative of a certain tradition in physics: the belief that if you could measure something precisely enough and model it rigorously enough, understanding would follow.

The work was rigorous and fascinating and, in a way that Dario would only fully appreciate later, deeply frustrating. The brain was unfathomably complex. A single cubic millimeter of cortical tissue contained tens of thousands of neurons, each with thousands of synaptic connections, operating according to principles that no one fully understood. You could spend a career studying one type of neuron in one region of one species and still not grasp the system's logic. Dario later reflected on this period through the lens of a physicist: he had come to the brain expecting the kind of clean, mathematical elegance that characterized the best work in physics. What he found was a system so complicated that the tools of physics could illuminate corners of it but not the whole.

The experience planted a seed that would germinate years later: the idea that biological systems, and biology as a whole, might be "too complicated for humans" to fully understand on their own. That understanding biology at the level of depth required to cure disease, extend life, and comprehend the brain might require a kind of intelligence that human scientists, working with human tools at human speed, could never achieve. He would return to that thought in 2024, when he sat down to write an essay about the future of AI and medicine that would become one of the most widely read documents in the field.

He earned his doctorate and was awarded the Hertz Foundation Doctoral Thesis Prize for his work. Then he moved to the Stanford University School of Medicine for a postdoctoral fellowship, where the focus shifted from pure neuroscience to applied machine learning: specifically, using early computational techniques to analyze biomedical data, including protein biomarkers for cancer. At Stanford, the two halves of his intellectual life, the physicist's love of scaling patterns and the biologist's immersion in the complexity of disease, began to converge.

• • •

One more thread runs through this chapter of Dario Amodei's life, and it is the hardest to write about, because he has spoken of it only rarely and with visible difficulty.

During Dario and Daniela's childhood, their father Riccardo became ill. The specifics of his condition have not been publicly disclosed in detail. What is known is that it was serious, that it was protracted, and that it cast a long shadow over the family. Riccardo grew sicker over the years. In 2006, when Dario was in his early twenties, his father died.

The loss was devastating in the ordinary way that losing a parent is always devastating. But it carried an additional dimension of anguish that would shape Dario's thinking for the rest of his life. Within four years of Riccardo's death, a medical breakthrough transformed the disease that had killed him. What had been roughly 50 percent fatal became 95 percent curable. Someone had worked on the cure, had pushed the science forward, had saved lives—but not fast enough to save Riccardo Amodei.

Dario has described this with a precision that suggests he has thought about it many, many times. The difference between his father most likely dying and most likely living came down to timing—a few years of scientific progress. Had the research moved a little faster, had the funding been a little larger, had the right minds converged on the problem a little sooner, his father might have survived. The margin between life and death was not a mystery of fate. It was a problem of speed.

For a person already inclined to believe that the world's hardest problems had objective solutions, this was a radicalizing experience. Understanding the universe was not enough. Fundamental scientific truth was not enough. The pace of understanding mattered. The velocity of discovery was, itself, a moral variable. Every year that a cure arrived late was a year of unnecessary death.

Dario has said that his father's illness was "one of the things that drove me" toward biology in the first place. But the death, and the cruel proximity of the cure, drove something deeper: a conviction that accelerating the pace of scientific progress was not merely desirable but urgent. That failing to do so carried a human cost that could be measured in the people who did not survive long enough to benefit from the next breakthrough.

"I get really angry," Dario would say years later, "when someone's like, 'This guy's a doomer. He wants to slow things down.'"

He did not want to slow things down. He wanted to speed them up. But he also wanted to make sure that, in the rush, nothing went catastrophically wrong. Holding both of those impulses in his head at once, the urgency to accelerate and the fear of what acceleration might unleash, would become the defining tension of his career.

• • •

By the time Dario finished his postdoc at Stanford, sometime around 2013 or 2014, the world of artificial intelligence was beginning to stir in ways that even insiders found hard to believe. A technique called deep learning, based on multilayered neural networks, a concept that had been around for decades but that had languished for lack of sufficient computing power, was suddenly producing results that shocked even its proponents. In 2012, a neural network built by Geoffrey Hinton's group at the University of Toronto had won the ImageNet competition, a benchmark for computer vision, by a margin so large that it effectively ended the debate about whether neural networks were a viable approach to AI. Andrew Ng, at Stanford and then at Google, was demonstrating that neural networks could learn to recognize cats in YouTube videos without being told what a cat was. At a time when most people in Silicon Valley were obsessing over social media and mobile apps, a small community of researchers—perhaps fifty people in the world, by Dario's estimate—was beginning to realize that something extraordinary was happening inside these networks.

For Dario, the timing was remarkable. He had spent years studying natural neural networks—the biological kind, made of axons and dendrites and neurotransmitters—evolved over hundreds of millions of years to convert sensory input into behavior. Now artificial neural networks, made of linear algebra and matrix multiplication and gradient descent, designed by humans and trained on graphics processing units originally built for video games, were starting to work. And they were working, he noticed, not because anyone had discovered the right algorithm or the right architecture. They were working because people had finally thrown enough computing power at them. The exponential curve that Kurzweil had pointed to was, it seemed, beginning to deliver on its promise. The crazy idea was turning out to be right, or at least, right enough.

It was time to stop studying the brain and start building the thing the brain had inspired.

In November 2014, Dario Amodei joined the Silicon Valley AI Lab at Baidu, the Chinese search giant, where a Stanford professor named Andrew Ng was assembling a small team to push the boundaries of what neural networks could do. Dario was thirty-one years old. He had a PhD from Princeton, a postdoc from Stanford, a Hertz Fellowship, and a Doctoral Thesis Prize. He had spent his entire adult life in academia. He had never worked in industry. He had never written production code for a commercial product. He had never managed a team, negotiated a contract, or worried about quarterly revenue.

And he arrived at Baidu with the slightly sheepish conviction that he was already too late. The deep learning revolution, as far as he could tell, had happened. Hinton, Ng, LeCun, Bengio: the pioneers had already claimed the territory. "My reaction at the time was, 'Oh my God, I'm so late to this area,'" he would recall. "This tiny community of fifty people—they're the giants of this field. It's too late to get in. If I rush in, maybe I can get some of the scraps." That was his mentality when he entered the field.

He was not too late. He was, almost absurdly early. The revolution he thought had already happened was, in reality, barely beginning. And the tool he would use to see what others could not, the simple, almost stupid act of making a neural network bigger and plotting what happened on a graph, was waiting for him at Baidu, in a room full of GPUs (graphics processing units) and a dataset of human speech. He just needed to start turning the dials.

# CHAPTER TWO

## The Bitter Lesson

*Baidu, Google Brain, and the Discovery That Scale Is All You Need*

The Baidu Silicon Valley AI Lab occupied a modest office in Sunnyvale, California, about forty miles south of San Francisco. Baidu Research was a satellite outpost of the Chinese search giant, run by Andrew Ng, a Stanford professor who had become one of the most prominent figures in machine learning. Ng had founded the Google Brain project before decamping for Baidu, and he had assembled a small team, perhaps two dozen researchers, to work on problems that required large-scale neural networks and abundant computing resources. The lab had a lot of GPUs. It had a lot of data. And it had a clear, specific goal: build the best speech recognition system in the world.

For Dario Amodei, arriving in November 2014, this combination of ingredients turned out to be a kind of intellectual gift. The task was concrete in a way that academic research often was not. At a university, a postdoc's job was to come up with a clever new idea, something novel, something that made your mark as an inventor. At Baidu, the job was simply to make the system better. He had a metric. He had data to train on. And he had GPUs to train with. The task was not proposing an elegant new algorithm. The task was concrete: what happens if you turn the dials?

“I kind of got lucky,” Dario has said, with a modesty that may or may not have been entirely genuine. “The task that was given to me and the other folks there—it was just to make the best speech recognition system that you can. A lot of data was available, a lot of GPUs were available. It posed the problem in a way that was amenable to discovering that scaling was a solution.”

He tried the simplest experiments. The speech recognition system was built on recurrent neural networks, RNNs, a type of architecture designed to handle sequential data like audio. Dario added more layers. He trained the models longer. He fed them more data. He watched what happened. And what happened was remarkably consistent: the models got better. Not in a jagged, unpredictable way, not with sudden breakthroughs interspersed with long plateaus. But smoothly and reliably, as though some invisible hand were adjusting a dial. More compute, more data, more parameters: better performance. Every time.

“I just tried the simplest experiments,” he recalled. “I was just fiddling with some dials. Try adding more layers to the RNN, try training it for longer—what happens? How long does it take to overfit? What if I add new data and repeat it less times? And I just saw these very consistent patterns.”

The work contributed to Deep Speech 2, an end-to-end neural network for speech recognition that dispensed with the hand-crafted feature engineering that had dominated the field for decades. The traditional approach to speech recognition was a Rube Goldberg machine of specialized components: acoustic models, language models, pronunciation dictionaries, hand-tuned post-processing steps. Each piece required years of expert knowledge to build and maintain. Deep Speech 2 greatly simplified this pipeline by training an end-to-end neural model from log-spectrogram features to character sequences, using a separate language model during decoding to improve accuracy. You fed it audio and transcriptions, and the network learned to do the heavy lifting that had previously required years of