"Understanding AI is a daunting task… This book offers a great starting point for those worried they'll never catch up."
— **REID HOFFMAN**, Cofounder of LinkedIn and Inflection AI

What You Need to Know About AI



A Primer on Being Human in an Artificially Intelligent World

JAMES WANG

"Understanding AI is a daunting task… This book offers a great starting point for those worried they'll never catch up."
— **REID HOFFMAN,** Cofounder of LinkedIn and Inflection AI

What You Need *to* Know About AI



A Primer on Being Human in an Artificially Intelligent World

JAMES WANG

PRAISE FOR *WHAT YOU NEED TO KNOW ABOUT AI*

"Understanding AI is a daunting task that becomes more difficult with the advances of each passing day. This book offers a great starting point for those worried they'll never catch up. James covers the essential history, principles, and implications of AI in a way that is accessible to neophytes but broad enough to offer useful insights even to AI professionals who often end up narrowly focused."

—Reid Hoffman, Cofounder of LinkedIn, Partner at Greylock Partners, and Coauthor of *New York Times* Bestsellers *The Startup of You* and *The Alliance*.

"A masterclass for understanding and investing in the most transformative technology of our time. From AI history to hardware and hype, this book cuts through the noise and delivers signal with clarity and insight."

—Rafic Makki, Head Technologist at Mubadala Ventures and former Executive Fellow at GlobalFoundries.

"This is the book I will give to everyone in my life to help them get to grips with the future."

—Jordan Schneider, Founder of ChinaTalk and Fellow at the Center for a New American Security (CNAS).

“James is one of the clearest thinkers in AI. This book is an encapsulation of everything that’s excellent about his work—deep systems-level thinking delivered through deceptively clear prose. *What You Need to Know About AI* doesn’t just tell you about AI; it gives you simple, powerful ways to understand its real impact on your job and the world. This is the most essential, grounded, and genuinely useful guide you will find on the most important technology of our time.”

—Devansh, Head of AI at Iqidis, writer of Substack bestselling newsletter Artificial Intelligence Made Simple.

“The quickest and clearest way I’ve seen to be well-informed on the most important topics in AI, and a book I will be sharing with friends who are trying to get up the curve on potentially the most important technology of our lifetimes.”

—Josh Blanchfield, Chief Investment Officer at Avos Capital Management, Former Senior Executive at Bridgewater Associates.

“James does a great job pulling back both the hype and horror around AI to paint a grounded but also encouraging picture of what the future can look like with this wide scope of exciting new technologies.”

—Dr. Joshua Reicher, CEO of IMVARIA, former clinical lead at Google’s medical AI group and faculty at Stanford Health care.

“A lucid, deeply engaging exploration that demystifies the most consequential technology of our time. James brings

together the rare trifecta of clarity, depth, and storytelling to deliver a guide on AI that is as thought-provoking as it is accessible. Essential reading for anyone who wants to understand what AI is and take an honest look at the implications of its integration into our lives. This book doesn't just inform; it empowers. It's the lighthouse for anyone navigating the fog of AI hype and fear."

—Brandon Reed, Founder of Dwellspring and Host of the *Ready, Set, Startup!* podcast.

"James is a polymath bridging AI history, cutting-edge algorithms, and the AI revolution with philosophical depth and venture capital savvy."

—Sergei Polevikov, founder of WellAI, writer of Substack bestselling newsletter AI Health Uncut, and author of Advancing AI in Health care: A Comprehensive Review of Best Practices.

"Really enjoyed this book. James always asks penetrating questions and that translates into an excellent synthesis of the pitfalls and promises of the AI revolution."

—Alex Campbell, Founder and CEO of Rose.ai and former head of commodities at Bridgewater Associates.

"I enjoyed reading this book by a successful and well-informed AI practitioner because it is easy to read yet captures some of the tough issues at the dawn of the age of AI."

—**Ashok Goel**, Professor of Computer Science at Georgia Institute of Technology and Executive Director of the National AI Institute for Adult Learning and Online Education.

Contents

1. Conclusion
2. Introduction
3. PART 1 - UNDERSTANDING AI
4. Chapter 1 - Luddites and Computers
5. Chapter 2 - Breaking Free from Science Fiction
6. Chapter 3 - Dueling Visions of AI
7. Chapter 4 - The Deep Learning Revolution
8. Chapter 5 - Demystifying the Magic of Modern AI
9. Chapter 6 - Know-Nothing Machines
10. PART 2 - AI IN THE WILD
11. Chapter 7 - AI as a Business Model
12. Chapter 8 - AI Hardware and Parallelism
13. Chapter 9 - AI at War
14. Chapter 10 - Why AI Won’t Kill Us—It’s Not Human
15. PART 3 - PUTTING IT ALL TOGETHER
16. Chapter 11 - The Missing Piece(s)
17. Chapter 12 - The New Luddites
18. Chapter 13 - Humans + AI
19. Conclusion - Peering into the Crystal Ball
20. Acknowledgments
21. Notes

Page List

1. 2
2. 3
3. 5

4. 6
5. 7
6. 9
7. 10
8. 11
9. 12
10. 13
11. 14
12. 15
13. 16
14. 17
15. 18
16. 19
17. 21
18. 23
19. 24
20. 25
21. 26
22. 27
23. 28
24. 29
25. 30
26. 31
27. 32
28. 33
29. 34
30. 35
31. 36
32. 37
33. 38
34. 39
35. 40

36. 41
37. 42
38. 43
39. 44
40. 45
41. 46
42. 47
43. 48
44. 49
45. 50
46. 51
47. 52
48. 53
49. 55
50. 56
51. 57
52. 58
53. 59
54. 60
55. 61
56. 62
57. 63
58. 64
59. 65
60. 66
61. 67
62. 68
63. 69
64. 70
65. 71
66. 72
67. 73

68. 74
69. 75
70. 76
71. 77
72. 78
73. 79
74. 80
75. 81
76. 82
77. 83
78. 84
79. 85
80. 86
81. 87
82. 88
83. 89
84. 90
85. 91
86. 92
87. 93
88. 94
89. 95
90. 96
91. 97
92. 98
93. 99
94. 100
95. 101
96. 102
97. 103
98. 105
99. 107

100. 108
101. 109
102. 110
103. 111
104. 112
105. 113
106. 114
107. 115
108. 116
109. 117
110. 118
111. 119
112. 120
113. 121
114. 122
115. 123
116. 124
117. 125
118. 127
119. 128
120. 129
121. 130
122. 131
123. 132
124. 133
125. 134
126. 135
127. 136
128. 137
129. 138
130. 139
131. 140

132. 141
133. 142
134. 143
135. 144
136. 145
137. 146
138. 147
139. 148
140. 149
141. 150
142. 151
143. 152
144. 153
145. 154
146. 155
147. 156
148. 157
149. 158
150. 159
151. 160
152. 161
153. 162
154. 163
155. 164
156. 165
157. 166
158. 167
159. 168
160. 169
161. 170
162. 171
163. 172

164. 173
165. 174
166. 175
167. 177
168. 179
169. 180
170. 181
171. 182
172. 183
173. 184
174. 185
175. 186
176. 187
177. 188
178. 189
179. 190
180. 191
181. 192
182. 193
183. 194
184. 195
185. 196
186. 197
187. 199
188. 200
189. 201
190. 202
191. 203
192. 204
193. 205
194. 206
195. 207

196. 208
197. 209
198. 210
199. 211
200. 212
201. 213
202. 214
203. 215
204. 216
205. 217
206. 218
207. 219
208. 221
209. 222
210. 223
211. 224
212. 225
213. 226
214. 227
215. 228
216. 229
217. 230
218. 231
219. 232
220. 233
221. 234
222. 235
223. 236
224. 237
225. 238
226. 239
227. 240

228. 241
229. 242
230. 243
231. 244
232. 245
233. 246
234. 247
235. 248
236. 249
237. 250
238. 251
239. 253
240. 254
241. 255
242. 256
243. 257
244. 259
245. 260
246. 261
247. 262
248. 263
249. 264
250. 265
251. 266
252. 267
253. 268
254. 269
255. 270
256. 271
257. 272
258. 273
259. 274

260. 275
261. 276
262. 277
263. 278
264. 279
265. 280
266. 281
267. 282
268. 283
269. 284
270. 285
271. 286
272. 287
273. 288

Landmarks

1. Table of Contents

WHAT YOU NEED TO KNOW ABOUT AI

A Primer on Being Human in an Artificially Intelligent World

James Wang



MANUSCRIPTS
PRESS

COPYRIGHT © 2025 JAMES WANG
All rights reserved.

The manufacturer's authorized representative in the EU for product safety is: Authorized representative Compliance Ltd, 71 Lower Baggot Street, Dublin D02 P593 Ireland (www.arccompliance.com)

WHAT YOU NEED TO KNOW ABOUT AI
A Primer on Being Human in an Artificially Intelligent World

ISBN
979-8-88926-582-5 *Paperback*
979-8-88926-583-2 *Hardcover*
979-8-88926-584-9 *eBook*

To humanity, with all our beautiful flaws and foibles.

CONTENTS

INTRODUCTION

PART 1
UNDERSTANDING AI

CHAPTER 1
LUDDITES AND COMPUTERS

CHAPTER 2
BREAKING FREE FROM SCIENCE FICTION

CHAPTER 3
DUELING VISIONS OF AI

CHAPTER 4
THE DEEP LEARNING REVOLUTION

CHAPTER 5
DEMYSTIFYING THE MAGIC OF MODERN AI

CHAPTER 6
KNOW-NOTHING MACHINES

PART 2
AI IN THE WILD

CHAPTER 7
AI AS A BUSINESS MODEL

CHAPTER 8
AI HARDWARE AND PARALLELISM

CHAPTER 9
AI AT WAR

CHAPTER 10
WHY AI WON'T KILL US—IT'S NOT HUMAN

PART 3
PUTTING IT ALL TOGETHER

CHAPTER 11
THE MISSING PIECE(S)

CHAPTER 12
THE NEW LUDDITES

CHAPTER 13
HUMANS + AI

CONCLUSION
PEERING INTO THE CRYSTAL BALL

ACKNOWLEDGMENTS

NOTES

“Progress is impossible without change; and those who cannot change their minds cannot change anything.”

—George Bernard Shaw

INTRODUCTION

If you lived from 1900 to 1930, the world completely changed. Clip-clops of horseshoes became the honks of automobiles. The flickers of oil lanterns became the clean, steady glow of electric lights. Mere dreams of flight became airplanes. Thirty years rendered the world unrecognizable.

The same span of time between 1980 and 2010 feels less significant. Perhaps that's just because many of us lived through it. Gradual change blinded us to how radically the world had already transformed. Counting minutes on payphones gave way to instant, constant communication with anyone in the world. Limited physical aisles in local stores gave way to ordering anything one could dream of, anywhere in the world, at the touch of a button. The act of paging through heavy, dusty volumes of encyclopedias has transformed into instant access to all human knowledge, which is now accessible anytime and anywhere. History books will almost certainly regard this revolution as a fundamental turning point. Now, with startling speed, we may be faced with yet another total shift in everything we know—artificial intelligence (AI).

Unfortunately, no one seems to agree on what it will do. Nobel Prize-winning computer scientist Geoff Hinton says AI has a 10 to 20 percent chance of destroying humanity in thirty years.[1] Famed venture capitalist and early web browser pioneer Marc Andreessen instead says AI will save humanity.[2] On its economic impact, another venture investor, Brad Gerstner, predicts that AI will be bigger than the internet and mobile—the technologies that drove our most recent thirty-year technological revolution.[3] In contrast, MIT economist Daron Acemoglu predicts in a Goldman Sachs report that AI would increase US productivity by a practically unnoticeable amount.[4]

If all of these experts disagree, what hope does a normal person have of understanding what's coming next?

To make matters worse, AI is not only moving extraordinarily quickly, but its most informative sources are also buried in technical jargon and niche publications. That leaves mainstream publications just as baffled, confused, and misinformed as most of the public. The public discourse remains a manic seesaw between glee and terror. What can AI actually do? What can it *not* do? What's real and what's hype?

This book is an attempt to answer these questions for everyday people who do not have a deep technical background and who are not following every new research publication. The beginning chapters build intuition around the aspects of modern AI: how technologies like AI impact society, how modern AI fits into its historical context, and what its technical underpinnings are. Next, the book walks through the practical aspects of AI, how it's used, and how it may reshape our environment and lives. Finally, the last chapters look at the relationship between humans and AI—why the rise of AI will likely increase the value of our very humanity…and save us from one of our greatest challenges in the coming decades.

While this book is designed to be accessible to those without a technical background, it does not sacrifice depth. I may not dive into every bit of technical jargon—I actively avoid most of it—but I try to build intuition on how AI actually works. The goal here is simplification without distortion. I aim to capture the essence of the technologies described in ways that are accurate and broadly applicable. As such, this book will remain useful even to those with some technical knowledge while those entirely without it may find a few chapters somewhat difficult but far from impenetrable.

I wrote this book because without a clear, *accurate* understanding of AI, it's impossible to decide whether we should fear it, ban it, welcome it, or worship it. If all we have is vague, disconnected headlines or misconceptions fueled by popular culture, we will be doomed to grope blindly while we stumble into the future. Many people have expressed both heady optimism and abject fear in response to ChatGPT and the subsequent, unrelenting avalanche of AI news. More and more people turned to me with rising excitement and terror. In response, I grew increasingly frustrated with the lack of a single, compact, nontechnical guide I could recommend—so I resolved to write it myself.

Why me? For many, AI entered their consciousness with ChatGPT's release in November 2022.[5] For me, it was already a decade-long fascination.

In 2013, I left the investment team of the world's largest macroeconomic hedge fund, Bridgewater Associates, on the East Coast to return to my childhood home of the San Francisco Bay Area. The reason was my excitement about what was happening in AI, especially with rapidly scaling computing power and computer vision hitting human levels of performance. During my homecoming, I cofounded a healthtech company utilizing AI, did a stint at Google's innovation and hardware division, and became a venture capital investor at a "deep tech" firm specializing in AI, robotics, synthetic biology, and other advanced technologies, where I am now.

I took a detour during college and my early career, but it isn't that surprising that I ended up back here. I've always been fascinated with technology—and I lived through the tectonic transition from 1980 onward that I described. One of my earliest memories was playing with the old green and black screen of a huge "laptop." I say "laptop" loosely. It would have crushed most laps. It also emitted a gentle hum reminiscent of a lawnmower or leaf blower. It only had a flashing box of a cursor, but I had fun moving it back and forth, hitting buttons that caused outraged bleats.

Later on, I remember the dial-up connection that would lock up our landline phones whenever we logged into AOL. And, even later, I remember the run-up of the internet boom, the bursting of the bubble, and the delayed but eventual triumph of internet start-ups like Google and Amazon, which became some of the largest companies in the world. I grew up in the epicenter of a changing world and can see it happening again.

A winding career with unexpected detours and careening turns has put me in a unique position. I'm a writer. I wrote anonymous web fiction read by thousands when I was a teenager and now have a popular Substack on AI read by investors, technologists, and policymakers around the world. I'm a technologist. I have done graduate work in machine learning with some of the greatest luminaries in the field and also built AI systems myself. I'm an investor and business leader. I've led teams, run companies, and served on boards of start-ups that have changed the face of the world. This

intersection of disciplines helps me understand AI's nature, its future, and how to communicate it clearly.

We particularly need such communication today. A 2024 YouGov poll of over a thousand Americans showed nearly 40 percent were somewhat to very concerned about AI ending humanity.[6] Fear of technology isn't new. And disruption of the status quo—and people's lives—has been happening long before even 1900.

AI, in a way, is different from everything that came before. Unlike the transition from shovel to bulldozer, landline to cell phone, or arrow to cruise missile, AI does something that no other technology has done before. It *thinks*. Or does it? After all, it certainly doesn't think like us. We have many popular media portrayals of AI in science fiction and movies, but few accurately portray the real-life technology that has already arrived. These technologies are no longer hypothetical. AI is here.

Will AI destroy us, as some fear, or save us, as others believe?

This book aims to provide a comprehensive understanding of AI's history and foundations, enabling you to form your personal opinions on this matter. If you're looking for a technical deep dive, this book isn't for you. We will talk about how AI works broadly, but we won't be talking about loss functions or tanh activations. Separately, if you are keen to learn how to use AI to boost your career and be "10X more productive," this book *also* isn't for you. Plenty of resources already exist that purport to do that—with varying degrees of accuracy and likely effectiveness.

This book is for those who want to learn about AI, its effects on them and their loved ones, and how to navigate our inevitable policy and social debates.

Before we get into the specifics of AI, we should first explore the broader patterns of technological change. AI may be unique in some ways, but many aspects of it *are* similar to other technologies. In the span of history and disruptions from new technology, we've been there before… and here we are again. Understanding that context is the only way to understand what is truly unprecedented about our current moment.

PART 1

UNDERSTANDING AI

CHAPTER ONE

LUDDITES AND COMPUTERS

Computers don’t exist anymore. Undoubtedly, that sentence sounds baffling, especially in a book about AI. But that’s because the *profession* of computer has been so thoroughly wiped out by *electronic computers.* Few even remember that it was once a title held by humans. During WWI and WWII, computers—mostly women, with men out on the war front—were critical for logistics, navigation, ballistics, and more. But by 1952, when the Association for Computing Machinery (ACM) started its now-famous journal, the job title “computer” had essentially disappeared.[1]

Liberated from pencil, paper, and human frailty, we’ve seen an exponential increase in both computational capacity and demand. Henry Ford’s Model-T, introduced in 1908 and a triumph of industrial prowess, barely had anything electronic at all.[2] In 2022, during the pandemic, Ford had over $3.1 billion in lost sales because it couldn’t manufacture enough cars. Why? Ford’s losses stemmed from a shortage of semiconductor chips, the digital brains that control almost everything in a modern car, from engine temperature to the oil change light.[3] We have put computational power—now being referred to more and more using the noun “compute”—everywhere.

This hasn’t caused mass unemployment. On the contrary, the United States in 1900, leading up to WWI and WWII, barely employed 4 percent of the population in technical or scientific fields of any kind. By 1950, when the computer profession had been wiped out, that proportion had increased to 8 percent.[4] By 2021, nearly a quarter of all jobs in the US were explicitly within science, technology, engineering, and mathematics (STEM).[5] Beyond explicit STEM jobs, nearly everyone on earth has access to computational capacity. The proprietor of your local corner store likely

utilizes digital spreadsheets for tracking expenses, surpassing the capabilities of any war planner from the 1930s. Even some of the poorest farmers today, like those I met during my time working with a nonprofit in Ghana, have flip phones with far more computational power than what NASA used to put a man on the moon in 1969.

OUR MODERN LIVES ARE MADE POSSIBLE BY TECHNOLOGY

This is literally how all technology has always worked. In the introduction, we described a changed world between 1900 and 1930. The cause, of course, was the tail end of the Second Industrial Revolution, which varied by country but was roughly 1870–1914.[6] Humankind started the period—outside of Britain, which transitioned somewhat earlier—as a species dedicating most of its population to farming to feed itself, with literal "cottage" industries and small-scale craftspeople.[7] The period ended with a version of our modern world that was dirtier, soot-coated, and more dangerous but broadly recognizable, featuring cities and factories that we'd be familiar with today. Manufacturing technology was a flywheel, allowing mass production of innovations like the threshing machine, which further increased farm productivity and freed even more farmers to become industrial laborers.[8]

Most of us would not enjoy living in that world. Workers toiled fourteen to sixteen hours, six days a week, with backbreaking labor. Child labor was common. Factory life was smoky, sooty, and extraordinarily dangerous. Overcrowded cities grappled with outbreaks of diseases like typhus and smallpox. At the same time, people still flocked to cities. While early industrial conditions horrify us now, people of that era were not abandoning comfortable lives. They were leaving similarly backbreaking labor on farms that tied them to their land, barely made them enough to survive, and had meaner conditions than almost any developing country today. City life, and the Industrial Revolution itself, allowed for cheaper basic goods—clothing and household items—and some level of wealth accumulation. It provided the dream of a better life.[9]

Objectively, we've seen a remarkable rise in material wealth and quality of life driven primarily by technology. Between 1900 and now, the world has seen global production value skyrocket from $4.8 trillion to $166.7 trillion (in 2021 US dollars), infant mortality go from common (around one in six children in 1900) to rare (less than 3.7 percent even including the poorest countries in 2022), and lifespan double from thirty-two years to seventy-three years.[10] It's impossible to deny the staggering progress we've made. We are the beneficiaries of multiple technological revolutions in the last century and a half. The Industrial Revolution. The Green Revolution. And, finally, the rise of the Information and Digital Age.

I think it's fair to say that the poorest in our society today enjoy better medicines, more varied foods, faster transit, more entertainment, and more basic comforts than the richest a century ago. The difference? Technology. Will the rise of the AI Age also herald incredible changes? We'll be covering this in-depth in later chapters, but our own technological transition is only just beginning.

Unfortunately, although "what comes next" has consistently been better than "what came before," that doesn't mean the transition won't be painful in the short and medium term.

LUDDITES OR COMPUTERS?

Luddite is now defined as "someone who is opposed or resistant to new technologies or technological change."[11] However, the term originates from a real movement in England during the 1800s. As previously mentioned, Britain was the first country to industrialize, doing so significantly earlier than the rest of the world and experiencing some of these transitional pains before others. Luddites were skilled artisans in textiles. They protested against the mechanization of their industry, fearing its threat to their livelihoods and craft.[12] Their story did not have a happy ending.

In 1811, workers stormed a factory in Nottingham to destroy textile machines. The government deployed soldiers who violently suppressed the uprising.[13] By 1813, Parliament passed laws making machine breaking punishable by death. Some Luddites were hanged. Perhaps worse, others were banished to Australia. The movement was dead.[14]

The Luddites were right, though. They knew their industry. And they knew what the technology did. As such, they knew it would degrade their wages, making it possible for their work to be done by unskilled workers and forcing them to take lower-paid work in factories.[15]

Technology may be a boon to overall welfare, but it isn't always good for everyone in the moment. Society was better off with more access to cheaper clothing and other textile goods. But the hard-earned skills of the Luddites were suddenly made far less valuable. Why did this happen? As mentioned at the beginning of this chapter, electronic computers wiped out human computers. However, STEM as a whole flourished. Anyone with an analytical mind undoubtedly found *more* opportunities with the mechanization of calculation. What made the difference?

For the Luddites, technology *down-skilled* their trade. Mechanization allowed far more people with less training to do what they could—at least close enough for most consumers. For computers, technology *elevated* their trade. Suddenly, the same numerically minded individuals *themselves* could do far higher-level tasks and spend less time on rote calculation. The difference isn't manual labor versus "knowledge work" either. Many manual professions—such as farming—have benefited enormously from technological progress. Workers ended up being able to do far more and improve their material well-being.

Farms went from nonstop, backbreaking labor that often left families barely able to feed themselves in the early 1900s to a forty-hour-a-week average job with wages above fifteen dollars per hour in 2023.[16] It's not glamorous, comfortable, or particularly lucrative, but it's a significant departure from the profession's beginnings. The farms also generate multiple orders of magnitude more produce.[17] Given that it's beyond the scope of this book to explore every instance of technological transition, readers will have to take my word for it that most technological transitions resemble computers rather than Luddites. That's scant comfort to the Luddites, though, and certainly no solace to those fearing job losses to AI. What if *their* field is closer to textile mechanization than computerization?

AI IS ALREADY CAUSING DISRUPTION

According to a survey by the Society of Authors, a quarter of illustrators and over a third of translators have already lost work due to AI. Furthermore, two-thirds of fiction writers and over half of nonfiction writers think AI will negatively impact their future earnings.[18] These worries have been backed up empirically. In a study using data from a leading online jobs platform, researchers found that substitutable skills like writing and translation saw a 20 to 50 percent relative decrease in demand since the introduction of ChatGPT. Interestingly, however, certain categories that utilized AI but weren't fully substituted, like "creative writing and explainer videos" and "lead generation," saw significant increases.[19] The story is obviously more complicated than simply writers becoming the new Luddites.

It makes sense that the emergence of Large Language Models (LLMs) would cause consternation, particularly among those operating in the same domain. Like a mechanical loom produces cloth, LLMs produce words. Will all writers and translators lose their jobs? Or what about programmers, who also have a profession composing words? At this point in the book, we don't yet have the tools to fully explore the question, but we will by the end of it.

LLMs and AI in general have too much fear, hype, and mysticism attached to them. We'll need far more information about the fundamental strengths and limits of AI, which will be the subject of the coming chapters. Only then can we answer the question—not only for writers, translators, and programmers but for everyone else too.

AI IS ALREADY SPREADING RAPIDLY

While writers worry about AI, other parts of the economy have enthusiastically adopted ChatGPT and other LLMs for everyday work tasks. According to a 2024 report by AIPRM, 75 percent of surveyed workers were using AI in the workplace.[20] Beyond that, familiarity with AI tools is high; a Pew Research Center survey found that 68 percent of US workers have some familiarity with AI tools. However, while many are optimistic, a significant portion (52 percent) are at least somewhat apprehensive about AI's impact on their jobs.[21]

Of course, this hesitation is completely understandable. We have already seen massive shifts in our economy, and this only adds to it. A study by the Brookings Institute found that more than 30 percent of all workers could see at least half of their occupation's tasks affected by generative AI. This impact is expected to accelerate trends such as automation in various industries.[22] While the outcome could mean workers transition to higher-paying roles, it's not inconceivable that certain jobs will simply disappear, and those workers might drop out of the workforce. Regardless of what happens, disruption is happening now.

OUR IMAGINATION OFTEN FAILS US

So what comes next after disruption? Historically, we have been terrible at figuring this out.

In the late 1700s, English economist Thomas Malthus observed a troubling trend. He saw the population explode in England with the rise in economic well-being. It seemed clear what would happen. He argued that humans would inevitably exponentially increase in population, but food supplies would never be able to keep up. Human hopes for "social happiness" were in vain. Essentially, humanity would be doomed to outstrip resources, and only war and famine could bring the population back in line. Famously, he was extraordinarily wrong, and improvements in agricultural yield (from technology) not only kept up with but easily outstripped population growth.[23] In November 2022, we exceeded eight billion people on Earth—a mind-bogglingly large number that Malthus could have never imagined. Maybe we can hope for "social happiness" after all.[24]

This pessimism in the face of disruptive change occurs in every era, though. In Jeremy Rifkin's 1995 classic, *The End of Work*, he calls the idea that we would replace jobs being destroyed by automation in the agricultural and manufacturing sectors with ones in science, engineering, and professional services a "pipe dream."[25] Of course, that is precisely what happened. STEM jobs are not neatly classified in US Bureau of Labor data, but "Professional and Business Services" skyrocketed (along with many other sectors) as manufacturing jobs precipitously declined in 2000.[26] This was really less automation and much more China manufacturing making its debut on the world stage, but regardless, we saw Rifkin's prophecy come

true with many jobs being destroyed—just by other humans abroad (which is a tougher issue).[27] Nevertheless, the economy rapidly shifted and created many jobs in other sectors.

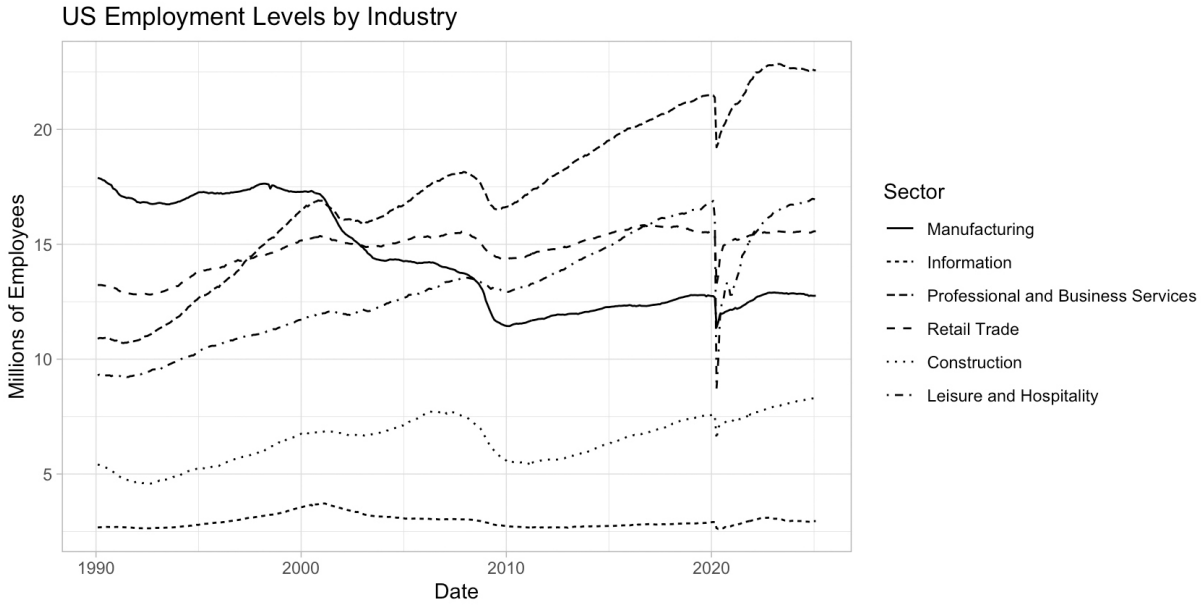


Image created by author using data from US Bureau of Labor and Statistics from 1990 to February 2025.

For better or worse, our complete inability to foresee the changes in our economies and societies in the face of disruption is a constant in human history. Malthus could see population growth, but he could not imagine how technology could increase material goods massively beyond anything he knew. Rifkin could see automation obsoleting jobs but could not see how the economy would retool itself.

Even just two years ago, I spoke with a successful businesswoman who channeled Malthus in chiding young people that they had to consume less, have fewer children, and live poorer lives because climate change is inevitable. She could see climate change but couldn't see that the trajectory of climate change had already been materially shifted due to solar improvements.[28] Technology causes disruption but often rides to the rescue as well.

Of course, today, we see the same worries from the previously cited surveys of authors, translators, and corporate employees. What comes next? Who would have predicted employment websites would have hundreds of job listings in California alone for "self-driving car engineer"?[29] While it's silly to try to predict exactly how our job landscape will shift, it's inevitable that it will happen.

But even if the economy shifts more rapidly than people expect, we'll still feel the pain of transition. As we showed in the study of the job listing site, in the short term, some people win, and some people lose. Sometimes people are computers. And sometimes people are Luddites.

In the case of the literal Luddites, as we discussed, vastly greater swaths of society enjoyed the benefits of greater material prosperity. Society wasn't going to return to the past. And given the collective increase in welfare, we wouldn't want it to. But the Luddites themselves were worse off, at least temporarily (not including those executed or sent to Australia, where conditions were unfortunately more permanent in nature).

It's impossible to know whether or not you are a computer or Luddite without understanding the technology itself. For Luddites, how the mechanical loom worked—and its threat—was clear. AI is more general purpose—and far harder to pin down. That's why we need to have a clear-eyed look at what AI actually is—and isn't.

CHAPTER TWO

BREAKING FREE FROM SCIENCE FICTION

From HAL 9000 to Skynet, we've long imagined our machines turning against us. These stories tell us something—not about AI but likely more about ourselves. Why is our popular culture and literature so rife with violent confrontation between humans and their artificial creations? If nothing else, we're always quite creative about it. Sometimes, AI inherits our hatred and fear, like Allied Mastercomputer (AM) from Harlan Ellison's classic "I Have No Mouth, and I Must Scream."[1] In others, we acknowledge AI's inhumanity and write them as coldly logical in their extermination of their creators, whether for self-preservation or some inscrutable purpose.

Hatred, fear, and even the drive for self-preservation are fundamentally *human* emotions. Our history is essentially one long tale of those emotions turned inward—driving violent confrontations that grew with us from small hunter-gatherer tribes to kingdoms and nation-states. Inevitably, even when we try to imagine nonhuman intelligences, it's difficult to break free of modeling them on what we know—ourselves.

Either way, we truly make AI in our own image. Existence leads to inevitable conflict. Consciousness is conflict. Stories to the contrary—where AI peacefully coexists with or even elevates humanity—are comparatively rare. They also happen to be some of my favorites, including Scottish author Iain Banks' *Culture* series and Isaac Asimov's *Robots* stories. I find them refreshing given the much larger corpus of stories where AI leads to humanity's doom.

I enjoy many genres of fiction, including fantasy and science fiction. As both an avid reader and an amateur fiction author in the past, I'm not one who thinks that only nonfiction is worth reading. Stories help us understand our humanity in familiar and unfamiliar settings. They help us wrangle with questions in our lives and appreciate the beauty of our existence—even in, and perhaps especially in, stories that are grim and harrowing.

Our imaginative instincts, while compelling, can be a dangerous mental model for the real world. Failing to distinguish fiction from reality is problematic, no matter the genre. In most cases, we would simply call it delusion. In the case of AI, some people call it prescient.

I've spoken with many "doomers"—people who believe AI poses an existential threat to humans—and most will cite science fiction at some point or another. Before signing his executive order on AI in 2023, President Joe Biden reportedly watched *Mission Impossible: Dead Reckoning*, where the antagonist is a malevolent AI and grew more worried about the technology.[2] With so little understanding of AI, the public leans on the imagination of writers—whether or not it's accurate.

MORE ALIEN THAN ALIENS

"It's hard to make things inhuman because you're always making things anthropomorphic, whether you like it or not. A human has no choice when they're writing," Tom Toner said to me on a video call. He is an English science fiction writer and the author of the *Amaranthine Spectrum*, which features AI as a core character. While human writers inevitably write AI with human characteristics, paradoxically the incomprehensible nature of AI's inhumanity is partly what makes it so fascinating as a literary device.

Even aliens, another common trope of science fiction, are a little more "human" and less often cast as an implacable enemy. Just look at popular recent Disney franchises in the Marvel or *Star Wars* series. We often fall into the trap of making aliens largely human in shape. More commonly than not, aliens share our distinct version of sexual dimorphism into male and female—just with different facial features, hair (or lack of it), and skin colors. But that just goes to show that we seem to empathize with the *concept* of aliens more in much of pop culture. All bets might be off if we

actually encountered life forms that don't just look like purple or green men and women with Hollywood good looks.

Nevertheless, alien biological life forms at least have some common bases with us. They almost certainly will have the fundamental drives for seeking nutrients, self-preservation, and reproduction. Survival pressures drive biological life toward convergent evolution, resulting in these shared experiences.

AI doesn't need any of that. Like modern seedless fruit, AI needs no innate drive for self-preservation, reproduction, or nutrition. It can be propagated endlessly by its creators. In that sense, it isn't like life. Because it *isn't* life. That's part of what makes it uncanny—both fascinating and frightening. As Tom Toner puts it, "There's the unknowable quantity of AI. It's more alien than alien."

A BAD MENTAL MODEL

Later in the book, we will discuss how AI might endanger humanity along with some of the best arguments for and against doomerism. To get the most out of this book, you must understand that science fiction is fiction. Imaginative depictions of other worlds and fascinating thought experiments of AI have their place. That place is not in deciding how you should account for AI's impact in your life or what you should do in public advocacy or voting. Fiction is fiction. Mistaking it for reality is delusion. This book's purpose is to help give a mental model based on how AI *actually* works, including knowing where it has come from, how it works now, and where it will go.

AI is a technology. It is more alien than alien in exactly the ways we have discussed, but so is a linear regression from your high school science textbook. It, too, has no recognizably human motivation and will senselessly and coldly calculate numbers. However, most students likely did not cower in abject fear of linear regressions—at least not out of literal fear for life and limb, anyway.

I don't say this to be blasé about AI. As we will see, it will likely have wide-ranging impacts on society, geopolitics, and even the physical

landscape of our world. Simultaneously, it isn't something that currently has either alien or non-alien motivations, and anthropomorphizing or catastrophizing what it can or can't do will inevitably lead to bad outcomes. In this case, a bad mental model is worse than no mental model. At least in the first case, we know what we don't know.

A REFLECTION OF US

After all of that preaching, I can't really deny that I personally was also drawn to AI from science fiction. There are simply too many themes that are beautifully explored through AI. Many of these are almost theological. We seem to love the theme of humans playing God and trying to create our own version of life—with our hubris causing our downfall. This is also a place where we can explore, letting our fantastical imaginations about what is possible run wild.

Tom Toner incorporated this theme in his own books. He mused, "For some reason, I keep including AI in my books, but… I don't do it on purpose." However, he then went on to describe some themes within his series, including ancient star-dwelling intelligences that have essentially solved all of the universe's mysteries—and started as AI. Few other beings could plausibly fill the same role. "That fascination goes all the way back to Mary Shelley, who wrote *Frankenstein* when she was nineteen… I suppose she would be the grandmother of science fiction." That was in the early 1800s. As Toner says, we've been fascinated with the concept of creating life—and having it go wrong—for a long time.

AI in science fiction is a reflection of who we are: our fears and psychoses, our hopes and dreams. Our fascination with AI may also be, as Toner quipped, "an inherent psychological fear of parenthood." I certainly understand *that* sentiment with a two-year-old daughter now. We've always used stories to cope with and comprehend difficult topics in our lives—from the oldest fairy tales told around a campfire to modern web fiction.

As a fellow lover of those stories, I don't ask the reader here to either discard them or disparage them. I merely ask that we set them aside for now. We are dealing with very *real* AI now, right in front of us. We will

want to stay in the realm of fact and not fiction. In our next chapter, I will also tell a story about AI. And this one is true.

CHAPTER THREE

DUELING VISIONS OF AI

One of the AI boom's unsettling questions is what happens to humans after their knowledge is "mined."[1] Would these people simply be discarded? The idea is disturbing. Alas, for better or worse, progress is inevitable.

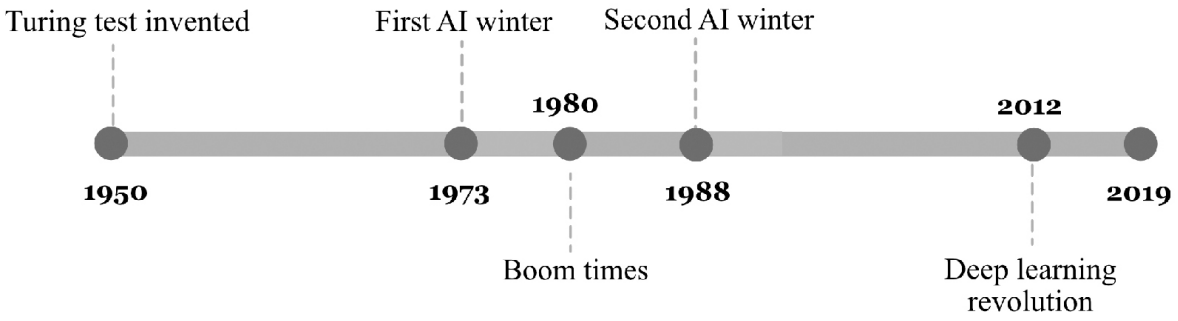
Another book on AI paints a picture of this frenzied climate. It references a national advertisement showing a computer and asks, "What if it could multiply thought as easily as it multiplies numbers?" Then it notes that these systems are promising a speed-up of professional work by a factor of ten—and sometimes twenty, thirty, and forty.[2]

With that kind of improvement, AI is undoubtedly going to be enormously profitable. In less than a decade, the AI industry has boomed from millions to billions of dollars.[3] Tracking one of the most popular AI conferences sponsored by the Association for the Advancement of Artificial Intelligence (AAAI), we see conference attendance almost double in three years.[4] What can possibly stop it? Well, AI could fall short of the lofty expectations heaped upon it.

After all, that's precisely what happened in 1988 when the AI bubble burst. The peak of the aforementioned AAAI conference was in 1986, with about fifty-three hundred attendees. Then it fell to less than two thousand in 1991. Finally, attendance languished at less than a thousand from 1995 onward—until recently, of course.[5] This generation of AI, which had widespread enthusiasm and investment, was characterized by expert systems, which we'll explore in detail.[6] Their fall helps put into context why the current generation of AI rose.

TWO DIFFERENT PHILOSOPHIES

Those outside of technology and AI are often unaware that there was a prior AI hype cycle. The cycle I described isn't even the first. In most classifications, it's actually the second AI boom, lasting from 1980 to roughly 1988.



A timeline of AI. Reproduced with permission from Sebastian Schuchmann.[7]

During the 1950s, *two* approaches had emerged for creating intelligent machines. In other words, artificial intelligence. These were "connectionism" and symbolic AI. Symbolic AI encapsulates *knowledge*—both through its internal representation and the carefully crafted rules that govern how it uses that knowledge.[8] In fact, symbolic AI is often used interchangeably with knowledge-based AI, with even the definitive textbook on AI by Russell and Norvig not making much effort to separate the two.[9] Symbols are manipulated, but knowledge is core—just read how a 1985 industry presentation described the implementation of these systems:

"The key factors that underly knowledge-based systems are knowledge acquisition, knowledge representation, and the application of large bodies of knowledge to the particular problem domain… don't tell the program what to do, tell it what to know… The task is elucidating and debugging knowledge, not writing and debugging a program."[10]

In a way, it has been the more "reputable" branch—so much so that this approach has also been known as "Good Old-Fashioned AI" or GOFAI, for short.[11] Symbolic AI is much more deliberately designed, with explicit knowledge coded into systems, and best known in the form of expert

systems—the very ones described in the hype cycle at the beginning of this chapter.[12]

This methodology is in stark contrast to the other school—connectionism. Marvin Minsky, known as one of connectionism's greatest critics, once scoffed that connectionists "take pride in not understanding how a network solves a problem."[13] Neuroscience serves as the inspiration for connectionism, which relies on distributed weights that, to humans, appear as a complete black box (as Minsky's criticism notes)—and which are most famously applied in neural networks.[14] Loss of transparency buys something else, though. Lack of explicit knowledge in the neural networks and lack of bespoke construction result in far more *flexibility.*

It boils down to a trade-off. Do you value knowledge or flexibility?

One can say that it would be beneficial to have both. However, at the time, these two approaches were often considered competitors and not complements—with proponents of each fighting for influence and research funding.[15]

AN EIGHTY-YEAR WAIT

The current generation of AI, which has captured the imagination of the world, is part of the deep learning revolution.[16] The best-known products of this revolution are ChatGPT and its competitors, which are all large language models (LLMs)—which are, in turn, neural networks.[17] These models are rooted not in the philosophy of the last AI boom, symbolic AI, but instead in the connectionist school.[18] As such, they can trace their lineage all the way back to the architecture proposed by McCulloch and Pitts for the artificial neuron in 1943.[19] In fact, as we'll see, one of the greatest champions of the connectionist approach in that era predicted that human-level communication was right around the corner.[20]

So, why did it take nearly eighty years to get ChatGPT in 2022?

As we will discuss in the coming chapters, AI can be broken down into compute, data, and models. At the dawn of the computing era and for many years after, neither sufficient computational power nor large enough volumes of data were available. For a model, the perceptron was limited, as

Minsky would point out. As one professor recalls, the original perceptron was just too simple. It had one layer while modern neural networks can have multiple layers and millions of nodes.[21] These multiple layers are the inspiration for the "deep" part of deep learning, which will be a topic we'll cover in detail, because as we'll see, modern AI is largely based on this. [22] On the other hand, while the models would be greatly refined over the years, including gaining many more layers, a lot of the basic concepts were already defined and likely could have been iterated upon. Yet the approach lost popularity and credibility for years. How and why did this happen?

As I mentioned earlier, this is a story of rivalry between two philosophies on how to approach AI. It's also, ironically, the story of two boys in Brooklyn.

TWO BOYS, TWO PHILOSOPHIES OF AI

Every research field is the collective work of innumerable scholars. AI is no exception. In a bit of cosmic coincidence, however, two of the best avatars of their respective schools of AI went to the same one in the 1940s. Frank Rosenblatt and Marvin Minsky both attended the Bronx High School of Science in New York City and were just a single grade apart.[23] In a way, the two were opposites not just in AI philosophy but in personality as well.

Frank Rosenblatt was a psychologist fascinated with biology. Back then, a defined area of "computer science" did not exist. It was the dawn of electronic computers, and those who studied it were yet to become "computer scientists." Still, you likely wouldn't have called Rosenblatt one even if you had the words. In fact, Rosenblatt only arrived at computers—and the artificial neural networks, perceptrons, which he became known for—as a way to better understand biology.[24] He exuded charisma and was larger than life. He was a "press agent's dream... a real medicine man..." described one annoyed scientific colleague.[25] Rosenblatt's sheer force of personality helped propel AI into the public's consciousness.

On the other hand, Marvin Minsky was a physicist. And while he also predated computer science, he lived to see it dawn as a field and became known as one of the best practitioners of it, especially within the field of AI. It isn't to say that he wasn't a warm and kind person. Speaking with

Istvan Berkeley, a professor of philosophy at the University of Louisiana at Lafayette who works on questions concerning the foundations of cognitive science, I was able to get some details as to who Minsky was. Professor Berkeley had extensive personal email correspondence with Minsky as a young scholar and shared some of those emails with me.

Berkeley's own recollections were of a man who, even as an academic titan at the time, was willing to generously spend time speaking to a relatively new academic. Stephen Wolfram recalls both Minsky's jokes and his "wonderful mixture of serious and quirky."[26] My impressions of those correspondences revealed a sober, cerebral man who naturally elaborated on complex topics for pages. He certainly was no showboat—unlike Rosenblatt.

Their clash may have been inevitable. Even Cornell University's profile on Rosenblatt, which was part AI history and part eulogy, had to set aside an explicit section on "Minsky versus Rosenblatt." The profile describes how they clashed during conferences, engaging in public debates that even now, decades later, young academics recall as "mind-boggling." Marvin Minsky, who outlived Rosenblatt by almost five decades, was described as the "loyal opposition" by Rosenblatt during his life. Although Rosenblatt came to it as a way to understand the workings of the brain, the perceptron was his passion, and he'd happily debate and promote it with all his considerable talents.[27]

It's tragic, in a way, that just two years before Rosenblatt's death, Minsky wrote the book that was widely credited with killing interest in Rosenblatt's beloved neural networks for decades.[28]

THE FIRST WAVE: ROSENBLATT AND THE PERCEPTRON (1958–1973)

In July 1958, reporters from major US newspapers crowded into a Navy research facility filled with blinking lights, hard clicks, dull motor whirrs, and occasional clunks from ingested punch cards. They witnessed the birth of something new, which journalists would literally call an "embryo" in their reporting. The Navy, and its charismatic spokesperson, described a fantastical vision for the room-sized machine. It had no limbs or mouth. But

soon, it would walk and talk. It would see and write. It would even become aware of its own existence.[29] It would be us, in manmade form—artificial intelligence. Or, at least, that was how it was described.

Electronic computers were still relatively new to the general public. And working with the US Office of Naval Research, Frank Rosenblatt had just unveiled a *learning machine*. Rosenblatt called it the perceptron, coined during his initial interview and in his paper released in the same year, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain."[30]

As I mentioned earlier, Rosenblatt was actually not the first to come up with the idea of an "artificial neuron." Warren McCulloch and Walter Pitts had proposed the theory as a model to simulate neurons in the brain in 1943.[31] Rosenblatt, however, added additional flexibility with his paper, and more importantly, actually implemented it in the real world with an IBM 704—a five-ton room-sized computer—which used its camera "eye" to distinguish cards with markings on the left versus those with markings on the right after fifty iterations.[32]

Even at this early stage, Rosenblatt made bold claims that perceptrons would be capable of human speech, communication and thinking.[33] Many would scoff at the pronouncements, but ultimately he was right. Telling the difference between left punches and right punches is no ChatGPT, but the successors of perceptrons did become LLMs, including ChatGPT.

Rosenblatt, the consummate salesman, described his perceptron as "the first machine which is capable of having an original idea." His sheer enthusiasm and charisma drew hundreds of thousands of dollars—millions in today's dollars—year-after-year in government funding, despite yielding no real results.[34] This obviously created enemies, with both resentment toward Rosenblatt and some believing that the entire enterprise of the perceptron was an egregious case of "overselling," which deserved to be taken down a peg or two.[35]

This made the environment ripe for a fall from grace with Marvin Minsky and Seymour Papert's seminal book in 1969 criticizing the limitations of

perceptrons—ironically named *Perceptrons: An Introduction to Computational Geometry*. A further tragic loss was Rosenblatt's death in a boating accident on his forty-third birthday in 1971.[36] Finally, the death blow came with the UK's Lighthill Report, which criticized both schools of AI for lacking any practical applications despite massive amounts of government largesse funneled into them.[37] This caused both schools of AI to undergo an AI "winter" in 1973 until the re-invigoration of the field in 1980—most notably, in symbolic AI, and expert systems in particular.[38]

THE SECOND WAVE: MINSKY AND SYMBOLIC AI (1980–1988)

Perceptrons was both a textbook and a critique. Most of the book is composed of chapters like "Algebraic Theory of Linear Parallel Predicates." The entire book is setting up for chapter 13, "Perceptions and Pattern Recognition," where the book switches to a questions and answer (Q&A) format where Minsky and Papert respond to a hypothetical questioner's defense of perceptrons—hammering home how poorly perceptrons would perform in a variety of problems.[39] The core weakness came from the simplicity I alluded to earlier. Minsky and Papert showed that a single-layer neural network could not learn XOR (exclusive-or) or any other "nonlinearly separable" function. In other words, they literally proved mathematically that the perceptron could never tackle anything of substantial complexity.[40]

What came after was almost a total shift in funding and research enthusiasm away from neural networks. Instead, the spotlight shifted to symbolic AI, the alternative favored by Minsky and his compatriots—and popularly cemented Minsky as the man who killed connectionism for years. This led to a "dry spell" for neural networks until 1980, when the work of scientists like John Hopfield (Nobel Prize winner in 2025 for his AI work in 1982) revived interest.[41]

However, even then, most of the commercial focus was on expert systems based on symbolic AI.[42]

THE RISE AND FALL OF EXPERT SYSTEMS

The perceptron machine may have been AI's debut into the public eye, but it and its related machines never became widespread. In the 1980s, that would no longer be true—but with symbolic AI in the form of expert systems. In the 1980s, over two-thirds of Fortune 1000 companies implemented initiatives to apply expert systems to daily business activities. The US and Japan also got themselves into an AI arms race. Japan announced the Fifth Generation Project, a decade-long plan to build intelligent machines, and the US countered with the Microelectronics and Technology Corporation (MCC). Companies like DuPont claimed that they built over a hundred expert systems that saved them over ten million dollars a year.[43]

One of the best-known expert systems was a research model called MYCIN. Developed at Stanford throughout the 1970s, MYCIN was used for diagnosing infectious disease and had a particular specialty in bacterial infections of the blood. Its demonstration of what was possible caused it to be widely studied and imitated.[44]

In a way, a chat with MYCIN would resemble what one might have with a modern LLM. MYCIN would go back and forth and ask details about a patient, inquire when samples were taken, and ask clarifying questions. Those questions, however, were formulaic and stilted. And underneath, MYCIN is completely different. MYCIN used rules and a context tree. It specifically needed to know an object, attribute, value, and level of certainty, which everything was translated into. Everything was either knowledge or rules to apply it.[45]

This has a huge benefit. The AI is completely auditable and transparent to a human. It's just rules and facts. While one can introduce probabilities, fuzzy logic, and other tools, it's still largely deterministic—meaning the same inputs will always yield the same outputs.

Determinism has another benefit. One of the biggest criticisms of LLMs has been their tendency to produce false or misleading information that looks factual—hallucinations.[46] As we'll discuss, LLMs, with their lineage largely stemming from connectionism, will never completely escape hallucinations. However, expert systems like MYCIN will *never* have them.

So what's wrong? A critical weakness of all expert systems—and symbolic AI in general—is that they tend to be quite rigid. Their internals look like math proofs. They are brittle when a novel situation that doesn't *quite* match their existing data banks comes along. Constantly searching for human experts to fill gaps is expensive. It's also a common occurrence when systems encounter the real world, which is messy and often doesn't match existing data, even when similar. These weaknesses of expert systems were widely acknowledged even at that time. They cannot deeply reason or learn from mistakes, and this limitation, combined with their inherently finite knowledge base, makes them quite narrow in scope.[47]

In a way, one can think of symbolic or knowledge-based AI as a student who memorizes facts. On the other hand, connectionist systems like neural networks resemble a student who forgets specific facts but remembers high-level concepts…and sometimes makes up facts. Each is useful in different situations, but in many messy real-world scenarios, the latter students will do better in more instances. But, in others, like medical diagnoses, you might prefer a student who only knows what's from the book and will otherwise say, "I don't know."

Ultimately, this is what killed expert systems. Ashok Goel—professor of computer science at the Georgia Institute of Technology, a leading expert in knowledge-based AI, and my professor on the topic—spoke at length with me over a video call on this point, especially relative to AI today.

"It was scale. A major thing that occurred was that they turned out not to be as scalable as people in knowledge-based AI had wished. A new problem, a new prompt, and you had to retrieve the closest matching case. The closest matching case in this particular study happens to be in the form of natural language, but it's still a case. What modern technologies, deep learning, and Large Language Model generative AI have done is scale up to a level that case-based vision and knowledge-based AI just did not succeed in doing."

It was scale…and from it, real-world flexibility and generalizability. This was already a strength of neural networks relative to expert systems. But it's also magnified by the fact that it's much easier to simply shovel massive amounts of largely indiscriminate and uncurated data into an LLM versus

trying to create a large, bespoke knowledge database and rules around it. This is somewhat of an oversimplification, as we'll discuss, but it's also not wrong.

The two metaphorical students have different strengths and weaknesses. But just because ChatGPT is confident, suave, and sounds like it knows what it's doing does not mean you necessarily want it diagnosing you or giving you legal advice—places where the precision truly matters.

SYNTHESIS

I think it's hard to deny that right now that connectionism is winning. Similar to when expert systems were triumphant, deep learning is ascendant. However, it's important to remember that each approach has strengths and weaknesses, and they do not have to be mutually exclusive.

On March 15, 2016, Google DeepMind's AlphaGo won a 4–1 victory over South Korean Go grandmaster Lee Sedol. Although AlphaGo's victory may appear similar to IBM's Deep Blue defeating chess champion Garry Kasparov in 1997, the two events are fundamentally different. AlphaGo's challenge is much harder. Relative to chess, Go has too many moves to "brute force" calculate the best next moves, which is how Deep Blue did it. The event caused an uproar, especially in East Asia.[48]

As a product of DeepMind, it's unsurprising that AlphaGo utilizes deep learning. However, that's not all it uses. Professor Goel explains that symbolic and connectionist AI camps often draw caricatures of each other to attack. He elaborates, "AlphaGo utilizes neural networks, which connectionists look at and call it their triumph. On the other hand, symbolists look at it and say it's all symbolic tree structures. Both sides say, 'I won.'" And, he says, they're both right. Both schools were utilized to make AlphaGo work as well as it did.

As we'll see shortly, while most advances are within deep learning, many concepts are already hybridized and can be divided into their respective schools about as well as you can unscramble an egg.

As for Minsky versus Rosenblatt, Minsky became a "pantomime villain, straight out of central casting," especially with the re-rise of connectionism

after its winter.[49] However, as Berkeley recounted to me, Minsky was a bit baffled that he and his coauthor, Papert, ended up becoming so hated in certain circles. In private correspondences Berkeley shared with me, Minsky praised certain aspects of Rosenblatt's work, including the learning function where the perceptron "learns only when it's made a mistake. That's the wonderful idea that's been ignored in the modern [neural network] era." He hardly sounded like a man trying to kill off another person's life's work. And at the end of the day, perhaps Minsky puts it best in one of his emails to Berkeley:

"In the field of theoretical physics, the researchers have a well-founded hope to discover a 'theory of everything' that has some simple and elegant form… But AI includes an 'everything' that is much more than what the physicists face…we shall have to learn how to combine and organize our ever-growing theories—and we ought to undertake that adventure with a positive view. Instead of envying the physicists, we could recognize that ours is the harder problem and one that in the end may turn out to be richer and more interesting."

History swung around from connectionism to symbolic AI and back. Predicting what comes next is difficult, but it's important to remember that the future may not be entirely so neatly split into two distinct schools. In fact, as we'll see, it's gotten quite muddled for some of the most cutting-edge models today. Nevertheless, it is still true that deep learning is ultimately connectionist in nature and has all of its strengths—and weaknesses. As we'll see, even those most intimately familiar with the research were shocked by how these manifested themselves with enough scale and data.

CHAPTER FOUR

THE DEEP LEARNING REVOLUTION

Our advisor gravely shook his head. “If you call it AI, it’s going to sound like a scam.”

The University of California Berkeley is one of the two academic powerhouses of the San Francisco Bay Area—often mislabeled Silicon Valley, even though that name technically applies only to a portion of the south area. Between 2013 and 2015, I took computer science and statistics courses from top AI professors while getting my MBA.

I returned from a career in East Coast finance to get closer to high tech and AI, and I intended to make the most of it. To that end, I tapped into as many start-up and entrepreneurship resources as I could on campus. It’s why our student team, despite pitching a half-baked idea we’d come up with barely half an hour earlier, had an advisor who had already exited one start-up and played key roles in others early on.

He said we couldn’t call the technology behind our idea “AI.” And I agreed.

Most people forget how stigmatized AI was between the early 2000s and 2010s. Industry leaders only began to embrace it in the mid- to late 2010s, and then it exploded into mainstream consciousness with ChatGPT 3.5 in November 2022. The line between AI and statistics is fuzzy at best. Linear regression, the same kind everyone learns about in high school statistics, is technically an “artificial intelligence” model. It’s a simple, early algorithm found in most textbooks, including in Russell-Norvig, the canonical introductory text of the field.[1]

Today, everyone clamors to claim their technology is AI. Those who do are obviously leaning into the hype, but with such a large gray zone, it's not always unwarranted. In contrast, in 2013, everyone shied away from the term even when it was warranted, preferring, "It's just statistics." Or, if one felt spicy, "machine learning." That statement is usually followed by, "It isn't AI."

Similar to the first AI winter stemming from the Lighthill report in 1973, the fallout was not confined to merely the school that flew too close to the sun. Symbolic AI, in the form of expert systems, dragged down the entire AI field's reputation with it.[2] But the smoldering coals of deep learning had already been laid in the 1980s—glowing quietly and waiting for the right spark.

That spark would come decades later with more powerful computers and more extensive data. This combination of compute, data, and models is a trifecta that will keep coming up. The concepts behind deep learning were not new. But deep learning was finally ready to ignite our current AI boom.

WHAT IS DEEP LEARNING?

Few names in deep learning are more esteemed than Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. This group is responsible for some of the greatest breakthroughs in the field. They have been fixtures of prestigious universities as well as Meta's and Google's AI labs.[3] As such, I will defer to them in defining deep learning, using their words from a 2015 review they coauthored in the journal *Nature* (emphasis mine):

Deep learning allows computational models that are composed of **multiple processing layers** to learn representations of data with multiple levels of abstraction… Deep learning discovers intricate structures in large data sets by using the **backpropagation** algorithm to indicate how a machine should change its internal parameters…[4]

As we had discussed in the last chapter, the major weakness of the original perceptron was that it had a single layer that made it useful for a relatively narrow class of problems.[5] Deep learning overcomes that challenge by adding multiple—sometimes even hundreds—of "hidden" layers in

between the input and output. Researchers have shown that even a *single* hidden layer gives neural networks the capability to approximate essentially any function—a property known as "universal approximation."[6]

This property certainly helps address one of Minsky and Papert's core criticisms of the perceptron. Berkeley points out that it's often overlooked that Minsky and Papert chose to base their proof on a neural network architecture that wasn't fully connected—a "limited order" constraint. This was deliberate. It reflected known biology. The field's collective decision to jettison the constraint presaged a more "if it works" era in machine learning.

Regardless, this universality was merely a theoretical result. It says that it's *possible* to approximate any function; it doesn't give any roadmap as to *how*. Specifically, additional layers increase computational cost but also raise a more profound issue: how exactly do you systematically update those internal layers to incorporate new data (a process known as "training the model")?

To illustrate this, assume that I tell you the answer to an equation is five. Let's also assume that we get this from adding three hidden numbers. What are the numbers? It could be one, one, and three—or two, two, one. Both are valid. And coming up with numbers that can work is quite easy. In a deep neural network, we are no longer dealing with simple addition. The operations could be complex, nonlinear, and across multiple layers that do *different* operations. For example, one could multiply the numbers, another could subtract a fixed value from all of them, yet another could cap values at a maximum number.

The classic "systems of equations" from high school algebra with linear operations is relatively easy to solve and find the solution(s). With a neural network, it's not clear which numbers, known as weights, we should nudge. A vast number of possibilities could fit the data—or, rather, approximate it since, in reality, statistical algorithms don't perfectly fit data. This means we're faced not only with an enormous number of possible solutions but also with no clear principle for deciding which are actually better for

improving the model. This ambiguity helped make multilayer networks an open research problem for years.[7]

This leads us to the second key characteristic of deep learning, which solves a problem that emerged from multiple layers—backpropagation.

Backpropagation works backward from the output, passing the "error" (difference between expected value and actual data) through each hidden layer in turn and propagating them throughout the neural network—hence the name. Specifically, it uses the chain rule from calculus to work through how errors should adjust weights. For our purposes, the precise mathematical details are unimportant, aside from the fact that backpropagation is both computationally tractable and a systematic, nonarbitrary process for training neural networks.[8]

Multiple researchers actually discovered this algorithm independently. Paul Werbos outlined it in his 1974 dissertation.[9] Later in the 1980s, Rumelhart, Hinton, and Williams rediscovered it, and their 1986 paper helped popularize and give backpropagation its name.[10] Hence, even while "connectionism" was in the wilderness, especially in industry, the groundwork was already laid for our current revolution. The outlines of our models were ready, but we still needed more. For most areas, we needed more data. For all areas, we needed more powerful computers.

Even in the mid-2010s, GPU-based computers had millions of times the computational power of machines in the 1990s. Today, they are more powerful still. Training multilayer models using backpropagation required computational power that vastly exceeded that available in the 1980s and early 1990s. It took many years before these algorithmic breakthroughs were combined with computers with large memory and computational speed to truly catch fire.[11]

IMAGENET AND ALEXNET

The slow, building simmer boiled over in 2012. That year, the ImageNet Large Scale Vision Recognition Challenge (ILSVRC), a competition that had become the de facto benchmark for progress in computer vision (and AI as a whole), saw a dramatic breakthrough. The task is simple for humans

and immensely difficult for machines: Identify the primary objects in an image from among a thousand possible categories, with thousands of labeled examples in the ImageNet dataset.[12] A new model called AlexNet slashed the top-five error rate from 25 percent to just 15 percent—a staggering ten-point drop that stunned research labs around the world. Unlike most of its state-of-the-art competitors, AlexNet used a convolutional neural network trained with backpropagation and accelerated by GPUs—an approach that, until then, had been largely dismissed as impractical.[13]

The result sent a shock through the AI field. Within months, nearly every major tech company and numerous start-ups were pouring resources into convolutional neural networks (CNNs)—the deep learning architecture AlexNet used—and deep learning generally. Almost overnight, CNNs became the dominant approach for almost all recognition and detection tasks.[14] Three years later, a Microsoft Research team announced a model with over a hundred layers and combined it with techniques to allow the training of such a deep network. As context, AlexNet had less than ten layers. This new Microsoft model had an error rate of 3.57 percent.[15] The estimated human-level performance for ImageNet is five percent.[16] In an incredibly short timespan, even by technology standards, AI went from stagnation to surpassing humans in a task utilizing one of our primary senses.

ALEXNET AND ME

This event had deep personal significance for me as well. I mentioned that my decision to come back to the Bay Area from East Coast finance was because of developments in computing power and AI. But the decision to leave for business school, made earlier in 2012, predated AlexNet. I was more excited about how machine learning was weaving its way into our everyday lives.

We had Google searches that felt like magic, bringing any information we wanted to us. We had Facebook and Instagram feeds that gave us an endless window to the world. We even had open machine learning competitions like the Netflix Prize, which had teams competing to best make movie recommendations for users on Netflix’s new streaming service.[17] Deep

learning was making a splash, even if the broader public hadn't felt it yet. As with many things in my life since, events came fast. Looking back, I can scarcely believe so much happened in such a compressed time frame.

In early 2013, I received acceptances from business schools and a prestigious writers' retreat for programmers (originally called Hacker School and later renamed Recurse Center), but I also received last-minute opportunities as well. My manager at Bridgewater offered me multiple times my already hefty hedge fund compensation to stay a few more months. MIT, my top choice, took me off their waitlist. Many of my friends were settling nearby in Boston. All signs—career, community, and opportunity—pointed toward staying on the East Coast.

It all came down to the night before the moving vans were coming to take my long-term girlfriend's (now wife) and my things to California. And I was running out of time to cancel them. Liz had gotten into prominent design and engineering graduate programs on the East Coast, but she was also struggling because she wanted to dive into start-ups in women's health *now*, not later. If we decided to stay, I'd have to literally wave off the moving truck pulling into my driveway in Connecticut.

The excitement about AlexNet was still percolating through tech-related IRC, Hacker News, and blogs. Bleary-eyed, sitting on the floor, I was making a pros and cons list on the only piece of furniture we hadn't packed, a whiteboard. I stopped to browse the internet, and I saw people talking about it. Then I read the research paper. AlexNet wasn't the only reason I made my decision, but when I happened upon it, I knew. This was it. It was happening. AI's breakthrough was here. So, Liz and I both took leaps into the unknown.

I turned down the offer from work, and we watched the movers load all our boxes onto the truck—including the whiteboard. We moved to New York City for the summer while I went to the Recurse Center and did projects in primitive game AI and web apps. I worked with ex-Google engineers and machine learning researchers. I even got a chance to pair program with Peter Norvig. Fast-forward again a few months toward the end of 2013, and I was sitting in the cramped conference room with Liz and a few other

group members at SkyDeck, the top floor of a Berkeley building and the space that would eventually become a prominent accelerator on campus.

We were pitching a women's health company that used AI, and our assigned advisor told us that we shouldn't call it AI.

LEGACY OF ALEXNET

I'm that annoying person at parties who corrects people when they say that AI's boom started in 2022 with ChatGPT. No, it actually started around 2012 and truly arrived around 2015 when AI reached human-level performance.

Vision has always held a special place in AI history and human imagination. It's arguably the primary way humans interpret the world. And it's incredible. We somehow turn vague signals from our retinas into *meaning*—effortlessly. We see how much this is all a mental phenomenon in experiments where subjects don inverted glasses and eventually see the world right-side up again after a few days.[18] This operation is one of enormous complexity, yet it occurs for us without a single conscious thought.

The fascination with this miracle of biology was why McCulloch and Pitts (original inventors of the neural network) did experiments with frog eyes and brains.[19] It's why Frank Rosenblatt's perceptron machine had an "eye" and did visual interpretation.[20] When AI reached the summit of visual recognition, it suddenly leaped forward in many different domains. Deep learning quickly saw breakthroughs follow in AI playing classic Atari games, machine translation, speech recognition, and many more AI applications.[21]

As for the authors of AlexNet, you've undoubtedly seen their work. We've met Geoffrey Hinton multiple times. He was one of the authors of that 1986 paper popularizing backpropagation. He was an author of the 2015 *Nature* review, helping define deep learning at the beginning of the chapter. He won the 2024 Nobel Prize in physics alongside John Hopfield for his work in AI.[22]

For the other two, Alex Krizhevsky (the "Alex") and Ilya Sutskever, they both went on to work on companies in deep learning. Krizhevsky joined Dessa, which aimed to apply deep learning to industry.[23] For his part, Sutskever went on to found a little company called OpenAI, which released a product called ChatGPT, whose technology shocked the world.[24]

GENERATIVE AI, TRANSFORMERS, AND LLMS

What exactly is ChatGPT, anyway? It's time to deep dive into the most famous example of modern AI that likely brought most readers to this book. Although I promised not to make this book too technical, we can't handwave *too* much because understanding what it is remains critical to understanding what it can do in the next chapter.

As I've explained, ChatGPT is a large language model (LLM). In particular, it is an LLM that uses the transformer neural network architecture.[25] *Every* major LLM uses it—including Google (Gemini), Meta (Llama), Anthropic (Claude), and even the Chinese company DeepSeek.[26] Under the hood, these model architectures are substantively the same, even if each has its own spin. For this reason, although I've used "ChatGPT" generically up to this point due to widespread recognition, I will largely be using "LLMs" from this point forward. Continuing the parade of terminology, LLMs are also in a class of algorithms known as "generative AI."[27] What does all of this mean?

Let's start with generative AI or generative models, which are interchangeable for our purposes, because it is both the broadest term and the easiest to grasp. Recently, the term has been genericized to mean any AI that generates content.[28] That's simple and descriptive enough, though not particularly informative. When I was taking graduate machine learning courses, generative models meant something very technically specific—in contrast to discriminative AI. Generative models learn the joint probability distribution while discriminative models learn the conditional probability directly.[29]

I'm going to simplify that using my own words because I suspect I lost many people in the last sentence. My explanation will lose a bit of technical precision, but I suspect it'll make more sense to most.

Most AI models operate on statistics. People generally think of discriminative models when they think of statistical models. For example, given a thirty-millimeter diameter, the fruit is most likely a strawberry (classification), or given a five-foot-nine male, the most likely American men's shoe size is 10.5 (regression). These models learn the association or mapping directly. In the case of generative models, they learn a probability *distribution.* Imagine your classic bell curve in high school but far more complex. It's not just the association; it's the whole "shape" of the data.

Using our prior example, a generative model learns what the entire distribution of fruits and diameters looks like, rather than jumping straight to linking fruits with diameters. Generative models can then use their learned distribution for classification or regression like a discriminative model—*or* generate new data points that look like the distribution. That latter part of the definition is the most relevant for how we think of generative AI today. It also highlights how these models work. They generate something that statistically *looks like* the data they were trained on.

As for large language models, the name is rather self-explanatory. You can use them for machine translation or autocomplete. Most are familiar with how autocompletion works on your phone. Suggestions pop up and you can simply keep accepting them. On most phones, once you get it started, you can keep hitting suggested words without typing a single thing. As I'll explain, an LLM is essentially a conversation where the model keeps accepting suggestions. We can and have implemented these models in ways that extend beyond just modern large language models and transformers.

One of my most popular GitHub (a public code-sharing site) repositories is one I published in 2014. It has been downloaded thousands of times and is a simple implementation of a Google paper, "Large Language Models in Machine Translation," with a funny name: Stupid Backoff. Just to be clear, the authors gave the model that name, and not me.[30] That system I implemented from the paper does not use transformers. It predated them. Transformers were only publicly unveiled in a 2017 paper called "Attention Is All You Need," again by Google.[31] As you might have noticed, many researchers take great pleasure in giving their models and papers clever

names. As you might have *also* noticed, Google created many of our core AI techniques over the years.

The reason why the paper is called “Attention Is All You Need” is because transformers use a mechanism called “attention.” Attention determines what each token (which are chunks of text—words, parts of words, and punctuation—but, for simplicity, you can think of them as words) “pays attention” to within input fed to the transformer. In the case of an LLM, it is the prompt provided by the user. Every token is simultaneously “paid attention” to with certain tokens weighted more heavily based on relevance. From there, the model uses a large number of matrix multiplications (from linear algebra) and then determines what the next token should be.[32] But why am I talking about this?

There are two significant takeaways.

First, this mechanism allows for a lot more parallelization of calculations. We will talk extensively in a later chapter about why parallelization is so important, not just for LLMs and AI but for modern computing as a whole. For now, the core insight is that, relative to older techniques, transformers allow LLMs to get much bigger because they can be trained in parallel—using all of the cores of a GPU, multiple computers, or even multiple data centers. Older techniques tend to require processing each token in sequence in order to derive meaning.[33]

This makes intuitive sense. The sequencing of a sentence matters, and you can’t just toss a random pile of words at someone without it becoming an unintelligible jumble. Transformers are different. Each token can “pay attention” to all other tokens at once. You don’t have to wait for each prior token to be processed before learning the association of subsequent ones. You can do them all at once with a bit of mathematical wizardry. How much larger did transformers get? Well, modern LLMs have basically ingested massive swaths of the internet and a huge number of books—sometimes in ways that raise serious legal and ethical questions.[34]

The second takeaway is that if we’re doing all this math to determine what the next token should be…that means, as I alluded to earlier, that LLMs are

just big fancy autocompletes. That should be simultaneously comforting—because that's the disappointing reality of AI behind the wizard's curtain—and awe-inspiring. Why the latter? Because if that's all they really are, what modern LLMs can do is truly remarkable.

CHAPTER FIVE

DEMYSTIFYING THE MAGIC OF MODERN AI

People use LLMs for all sorts of things. One person I know made a pros and cons list about her relationship to decide whether to leave or take the next step. Another used it to generate an analysis of his company's value and sent that off to an interested public company acquirer for a deal potentially worth millions of dollars. My wife and I used it to pull studies and analyze the probability distribution of IVF—with complications from a sex-linked genetic condition—to estimate how many eggs we'd need to retrieve for one successful pregnancy.

That's an awful lot of trust placed in what is, at its core, a glorified autocomplete. Yet these examples are all reasonable because they share a single, crucial characteristic that we'll return to in the next chapter.

For some of you, though, a mysterious "crucial characteristic" to use LLMs like this probably feels superfluous. Some freely use LLMs for these kinds of things and more. It's reasonable—mundane, even. For many, they've become a part of daily life. As of early 2025, ChatGPT has four hundred million weekly active users, and the company is targeting over a billion by the end of the year.[1]

When I was growing up, if we were dubious about something our peers said, we just had to decide whether we believed them. We certainly weren't trekking to a library to pull out an encyclopedia—let alone booting up a home desktop computer, dialing up to the internet (a screeching, hissing affair that kids today will never know), and going to a search engine like

Ask Jeeves, Lycos, or Yahoo. The generation after me just whipped out their smartphones and Googled it. Kids today? They ask ChatGPT.

But have you ever asked an LLM how many Rs are in “strawberry”? A viral Instagram reel by tech comedian Alberta (@alberta.tech), with hundreds of thousands of likes and millions of views, showed major LLMs —including ChatGPT, Llama, and Claude—all confidently getting it wrong. They say two. The correct answer is three. You can count it yourself. Claude even argues with her when corrected, saying it “respectfully disagrees.” Why? Alberta then breaks down why this happens—a process we already touched on at the end of the prior chapter. It’s because transformer-based language models don’t actually “see” any words.[2]

As a recap, LLMs process text by breaking it into tokens and then use that “attention” mechanism we described to represent each token as a numerical “vector”—a bundle of numbers. Those vectors are passed through a stack of computations applied to predict the next token. The model *literally* never sees “strawberry,” so it cannot count letters, even if it wanted to. It’s not even working with a number that means “strawberry.” It’s working with a vector that represents “strawberry” *and* the context of all the other tokens around it. In other words, “strawberry” surrounded by different words? That’s a different vector.

I wouldn’t try this test anymore, by the way. LLMs update frequently, and developers quickly patch cracks in the illusion of infallibility. But there have been, are, and will always be moments when we glimpse what these models truly are. That is, huge statistical machines that autocomplete words —including their own output—to build responses that sound incredibly human. But they *aren’t* human, even if, as we covered before, they are remarkably easy to anthropomorphize. That just makes it even *more* incredible that they do what they do.

IF YOU CAN’T TELL, DOES IT MATTER?

Calling LLMs fancy autocomplete algorithms is dismissive and overly reductive. It’s also still true. The fact that LLMs can follow instructions, edit code, create analyses, and even *attempt* to count Rs in “strawberry” is unbelievable when you know what is under the hood. The mechanism may

be predicting one single token at a time, but if it perfectly emulates what a person would write—a doctor for a diagnosis, an editor for a review, or a programmer for a feature—does it really matter?

In the 1980s, American philosopher John Searle posed a deceptively simple but brain-bending philosophical question known as “The Chinese Room.”

Imagine being locked in a room as an English speaker who knows no Chinese. Instead, you’ve been given a bunch of protocol books and instructions on how to respond to specific unintelligible symbols (Chinese characters). Someone keeps passing you notes in Chinese, and you respond. You have no idea what’s going on, but you keep passing notes. You convince the Chinese-speaking person outside that you, indeed, are human.

That’s the basis of the classic Turing test for AI. When an AI can’t be seen and is hidden in a black box, can it be indistinguishable from a human? In this Chinese Room thought experiment, you pass the Turing test, but you’re not a computer or a black box. *You* are inside the black box, and you *know* that you do not understand anything. You don’t know the language. You are just mechanically following instructions. But to the person outside, you seem fluent. Searle argues that this is analogous to what computers do—manipulate symbols with no understanding.[3] Intuitively, when put this way, we seem to understand something is missing.

Fortunately, this is still a somewhat theoretical question. As we will discuss in the next chapter, an LLM does *not* perfectly emulate a human for any of the expert tasks we gave at the beginning of this section. And even for their core function, written communication, LLMs still don’t reliably pass as human. A Temple University study in 2024 found that human evaluators could correctly distinguish AI from human-written texts 57 percent of the time. Not a great rate of fooling humans, though far from comforting if someone were accused of using AI, as many students have been.

Domain familiarity matters, though. In this generalist pool, which likely spends more time reading social media than general-interest texts (unlike my readers here, of course), participants were 78 percent accurate in uncovering AI in social media comments.[4] What if we have a population

that has expertise in both LLMs and text? A 2025 study of writers who frequently used AI for writing tasks found that a group of these experts trying to sort between AI-written and human-written articles mislabeled only one out of *three hundred*.[5]

I assume, however, that this will not last. Just as LLMs no longer have an Achilles heel in the form of the word "strawberry," model companies will rid LLMs of their most telltale writing tics. It will become harder and harder to distinguish if you're talking to an AI or a human. If so, does that mean we've "made it" to human-level AI—often vaguely termed artificial general intelligence (AGI)?[6] Is it so close that it doesn't even matter?

No. And this matters more than mere philosophy because even if humans can be fooled by LLMs in a casual conversation, that doesn't mean they think like a human. That matters because our ability to think goes beyond mere conversation, and that dictates the strengths and weaknesses that are evident in LLMs and all other forms of deep learning.

As Gary Marcus aptly puts it, "The Turing Test is a test of human gullibility, not a test of intelligence."[7]

NOT YET HUMAN

Humans do more than merely communicate. Our mental life is deeper than that. Language itself, however, feels special. The ability to convey complex, abstract ideas with it separates us from other animals. That's likely why an LLM like ChatGPT broke through to the public rather than other achievements in deep learning. However, even if LLMs do perfectly emulate humans in this domain, that doesn't mean they are approaching any reasonable definition of AGI.

Yann LeCun, in particular, likes to describe current AI as dumber than a housecat. Cats, after all, can remember things, reason, and plan—even if the task is as simple as remembering you provide food, you will continue to provide food in the future, and they can sit around and meow to get it. No current AI can even begin to claim that.[8] Now, LLM model companies *are* trying to address that. In 2024, OpenAI introduced a "memories" feature in

ChatGPT to reference some past conversations.[9] This feature was updated in mid-2025 to remember all your conversations.[10]

Does that make ChatGPT closer to AGI? It's a step in the right direction, but "that's still a cheap trick" was the answer I got through our, at times, choppy connection. While he was on a research trip in Spain, I spoke over video call to Alejandro Piad Morffis, professor of computer science at the University of Havana, who specializes in research on language models.

He holds a similar view to Yann LeCun and dismisses current efforts like ChatGPT memories as insufficient. He notes, "I'm not anti-LLM or anything. On the contrary, all of my research is on language models. My day job and pay depend on language models working." Beyond that, he's a computationalist; hence, the name of his popular newsletter, "The Computist Journal."

That means he doesn't subscribe to the theory of duality in the philosophy of mind, which argues our minds have something beyond the measurably physical—one might say, a soul. Given that he isn't a dualist, he does not believe anything prevents computers from reaching AGI if we actually approached it correctly. As for why our current AI isn't there, he describes multiple challenges.

One challenge is in capabilities. A human (or cat, as per LeCun) can go out into the world, experience it, and continually learn things. An LLM cannot. Morffis's own research involves this, experimenting with how far an AI can get if it is provided the ability to remember (in a more robust way than ChatGPT's simple memories feature) and have access to various computational and perceptual tools. Although language is powerful, he argues that one cannot understand many underlying concepts purely through language. As he says, "You cannot learn physics by reading about physics. You have to perform experiments in the world to understand how it works."

My own analogy to this is one cannot *truly* explain color to someone who is blind. Something would still be missing with just words to describe it. How much worse is this for a machine that has access to *nothing* but language?

And, even worse, nothing but the *specific* set of language fed to it? It can describe everything. It has also experienced utterly nothing of what it describes.

In a 2021 paper, Goel makes a similar point to Morffis. Cognitive science has made extraordinary progress in placing "significant parts of mind outside an individual human's head. However, the same kind of expansion of scope has not yet occurred in AI." He explains that new theories of socially situated intelligence, embodied intelligence, and more could significantly add to our understanding of intelligence's true nature.[11] Something in the broader *environmental* context makes our intelligence what it is.

Could we solve this by giving AI bodies like humans (or cats)? Would that one step suddenly leap them into becoming AGI? Morffis argues no.

The other challenge lies in the fundamental nature of connectionism and deep learning. LLMs and similar models are merely learning statistical distributions, which is not how our minds work. How do we know? Morffis describes how children learn language, which is by "sampling" something on the order of hundreds of thousands of sentences before they start talking. That scale of data does not contain enough semantic information to perfectly learn English grammar. A language model trained from scratch like that would be completely useless.

Deep learning requires millions of examples. Kids require just a few. That suggests something deeper than mere pattern recognition. Something is biologically ingrained beyond neurons strung together. It's something like knowledge structures—as we've described in symbolic AI. Noam Chomsky's theory of "universal grammar," widely accepted by biologists and linguists, hypothesizes exactly this.[12]

My daughter, who was around one year old when I started writing the first draft of this book, could recognize—and assertively yell about—"doggie" after a single instance of seeing a dog. While there were some misclassifications of "doggie," like with other four-legged animals, she got the general concept (and corrections) immediately. No deep learning system

today can accomplish such a feat. To achieve its world-class performance in 2012, AlexNet required millions of images in its training set.[13]

François Chollet, a prominent AI researcher, proposed a framework for intelligence while at Google in 2019: "The intelligence of a system is a measure of its skill-acquisition efficiency over a scope of tasks, with respect to priors, experience, and generalization difficulty."[14] As Morffis explains, LLMs are good at generalization—which, as I will explain, is part of their "superpower"—but their skill-acquisition efficiency and ability to adapt to totally novel situations are terrible. That can be fine for creating AIs or AI-powered robots that have high economic value. These are "philosophical zombies," as Morffis describes them, but they won't truly be intelligent—not in the way that humans or even animals are.

REASONING WITHOUT REASON

Overcoming that barrier, supposedly, is exactly how OpenAI's o3 burst onto the scene in late December 2024. Chollet was the main architect of a test known as ARC-AGI (Abstraction and Reasoning Corpus-AGI) based on the principles of his framework—meaning a test that explicitly aimed to capture Chollet's notion of intelligence. OpenAI's GPT-3 started in 2020 at a score of zero percent. Even the powerful GPT-4o scored a mere five percent. Many other state-of-the-art models fared no better. OpenAI's second iteration of its "reasoning" model, o3, blew the doors off with a score of 87.5 percent.[15]

Google researchers had noted before in 2022 that using "chain of thought" by walking an LLM through intermediate steps of a problem tended to improve its performance on complex problems.[16] While this was interesting, it's not ideal. At the time, I thought it was mainly a "nifty" technique. With a human-in-the-loop, this is both laborious and could also allow for leaking information into the model with leading questions. OpenAI's o1, their first reasoning model, changed this in mid-2024. o1 prompted itself and allowed for long chains of "internal conversation" until it reached an answer.[17]

This development has huge business implications, as I wrote and noted at the time. Effectively, because one can have "intermediate" tokens that are used to get a better final answer, one can add more compute time—in other

words, money—in order to get better answers. This means that the potential compute cost (and, ultimately, power cost, as we'll discuss) for useful AI could skyrocket.[18]

o3's accomplishment was incredible, as was the cost. Initially, reports had cited that it cost over a thousand dollars per task. Months later, in April 2025, the estimate had crept up to thirty thousand dollars per task.[19] Again, this has huge business implications, but the question is, does this change the nature of LLMs from being the association machines—fancy autocompletes—that they have been described as so far?

As I described at the time, reasoning substantially increased performance without significant changes in base architecture and is (maybe) auditable by reading its intermediate outputs. But the way it functions is still like that of a statistical model. My description, then and now, is that LLMs will typically give you the most probable autocompletion. If you prompt it better or use reasoning, it will "wander" within its vector space inside of its neural network to a more precise location that better fits your query.[20]

Another way I like to describe LLM reasoning is the models feed themselves. In other words, the model's *own output* becomes part of its *input*. This approach can allow the model to refine its output. But think about it—if what's helping the model go further is *still* from its internal statistical distribution, it will never go somewhere totally new.

A good example comes from Marc Andreessen and Ben Horowitz on the *a16z* podcast. They noted that prompting an LLM for "secure code" often yields better programming output. Why? Because it shifts the model's sampling distribution toward high-quality examples. A security-conscious coder responding or asking questions on Stack Overflow or other online forums is generally better than the mediocre average.[21] This still doesn't give you true adaptability outside of a training set.

As Google machine learning engineer Logan Thorneloe commented on people's surprise that LLM reasoning had limits in new tasks or environments:

“I’m also intrigued by how many people are surprised by the fact that models can’t perform outside of their dataset when by definition this is how deep learning works. The real way to get them to generalize further is by finding unique ways to expose [them] to more data.”[22]

Inherently, reasoning models are still deep learning models—with all of their strengths and limitations.

Despite achieving a high score *overall* on ARC-AGI, Morffis noted in our discussion—as I did in a piece I wrote at the time—that o3 failed certain simple visual puzzles in the test that even children would be able to solve. The task required too many levels of abstraction, even if they were simple ones that humans largely find trivial. AI, in contrast, found it extraordinarily difficult.[23]

ARC Prize, the organization in charge of running the ARC-AGI benchmark, isn’t standing still. They recently released ARC-AGI-2, where humans still score the same, but modern LLMs score zero, and even reasoning systems score in the single digits.[24] It essentially doubles down on what we’ve already seen modern AI fail to handle well.

Even so, I expect this benchmark will likely get “solved” before long as well. One part of this is future-proofing my book. If the ARC-AGI-2 is “solved” immediately after this book’s publication, I’ve at least called out the possibility! More seriously, Morffis pointed out during our discussion that the AI research community often over-indexes on benchmarks. As such, benchmarks shape behavior. Models, like students cramming for standardized tests, can ace them without actually learning anything general.

We’re still missing something. Morffis put it: LLMs optimize for plausibility and not correctness. Unlike the expert networks of the past, they can generalize to novel problems. But also unlike expert networks of the past, there is no guarantee of the answer being right—and there never will be.

EVERYTHING IS A HALLUCINATION

On November 1, 2024, communications professor Jeff Hancock of Stanford University filed an original expert declaration for a serious court case

regarding Minnesota's ban on deepfakes (falsified AI images and videos) used to influence elections. On November 27, he filed an amendment, stating he "did not intend to mislead the Court or counsel." Hancock had used fake citations in his original declaration because he used ChatGPT. He had fallen victim to LLM hallucinations.[25]

The court case's subject—AI misinformation—made this deeply ironic. But Hancock was an expert that should have known better. He is codirector of the Stanford Cyber Policy Center and the founding editor of the *Journal of Online Trust and Safety*, a leading journal that covers misuse of the internet and AI.[26] It's like the head of IT security clicking on a suspicious phishing link.

Lawyers keep getting themselves in trouble using cases that sound real but aren't real.[27] Again, as Morffis described, LLMs are great at *plausibility*, not *correctness*.

That really should give anyone who regularly uses LLMs pause. Not only are LLMs able to give false information, but they are incredibly convincing because, in a way, that's exactly what they are designed to do. They learn a statistical distribution and draw from it. To do so correctly, the information produced by the model is going to closely resemble *what is found in the distribution*. The trouble is that the real world has facts that aren't quite so probabilistically mutable.

Just imagine what might happen if you consult an LLM on a critical medical issue. It is usually accurate because many medical issues are reasonably clear-cut, which you'll hear more on later. But right or wrong, it will *sound* right because it is optimized to sound like the distribution. Having a *plausible* conversation rather than a *right* one is problematic when you have high stakes and aren't consulting other sources.

We've started to call these kinds of mishaps hallucinations—formally defined as LLMs outputting false or misleading information that looks factual.[28] Many have labeled hallucinations one of the most pressing problems in AI. I've personally sat in dozens of pitches from AI companies declaring that fact. Coincidentally and fortuitously, I happen to be sitting in

front of a company that has solved *exactly that*. They have significantly reduced or even *totally eliminated* hallucinations. What are the odds? As it turns out, really high whenever someone is pitching the idea.

I'm quite forgiving of bad pitches. I'm not forgiving at all with technical bullshit. These companies typically fall under the latter category.

Now, it is certainly possible to have LLMs output objectively accurate information *more often*. We'll even cover techniques to do so later. But remember how deep learning is different from expert systems. It's *flexibility* and *generalizability*. That core strength makes all of its branches so useful—from LLMs handling novel sentences to computer vision algorithms tackling new pictures of cats. What is the inverse? What is the critical weakness of models like LLMs that is the core strength of expert systems? *Knowledge*.

Remember my analogy of the two students. The first memorizes information and can recite facts but cannot apply it easily to new situations. The second generally knows concepts but can't really rattle off any single precise fact. An LLM, which is based on deep learning, is the second type of learner. The fact that "everyday" questions come back correctly is largely a function of people on the internet asking the same kinds of things over and over again. This means that most questions draw from the "center of the distribution." Imagine a classic bell curve. Most people simply draw from the center and almost always get the right thing back. Think of it this way. No user follows the exact same path, but on a well-trodden trail, it's hard to get lost.

That assumption immediately goes away when either the question veers to something more precise, where a "gist" answer doesn't cut it, or the situation has high stakes. Law keeps getting itself in trouble because it's *both*.

That does not mean that deep learning or LLMs are not extraordinarily useful. That mystery characteristic I mentioned at the beginning of the chapter that made my instances of "high stakes" cases safe? It's an oracle. We'll get into that immediately in the next chapter, so hold that thought.

Before we get there, why *can't* we fix LLMs?

Pedantic questions like "What is truth?" are often asked in pungent college dorms. While it's not a completely unwarranted question, in practice, we have certain kinds of factual mistakes we want LLMs to avoid. Some things are colored in gray, but there are black and white errors in hard facts that can be falsified by reliable sources.

An obvious first step would be to force reliance on *only* trustworthy sources. Similar to how we chide high school or college students not to use unreliable sources, we'd want to ban Wikipedia, sketchy sites, and anything else that isn't within a "blessed canon." Someone—not some machine—would need to make that determination. Beyond that, we can't simply have the LLM "get the gist" of the canon and potentially make things up. We'd need it to follow a more reliable system—say, one based on logic—to derive generalizations or adaptations of its facts. In other words, we need to reinvent symbolic systems.

That's why we can't fix them. We can "hybridize" systems to some degree, but it's *part of the nature* of LLMs to do this. We simply have to accept it and utilize it *despite* knowing they will get it wrong sometimes.

More profoundly, realize that *we* see hallucinations. We have some kind of external verification of truth that tells us some outputs from an LLM are true and others are false. An LLM does *not*. For an LLM, everything is the same—just tokens in a sequence. There's no difference between "true" tokens and "false" tokens. In a sense, *everything* is a hallucination for an LLM. It merely predicts what probabilistically comes next. That's why Morffis has often used the phrase first coined by Emily Bender and her coauthors in a 2021 paper: "Stochastic Parrots." It sounds like it knows things, but it never will.[29] The world is merely a fever dream of spitting out the next token.

Nevertheless, even if LLMs—and other deep learning algorithms—don't *actually* know anything, that doesn't mean they aren't useful. It doesn't even mean they can't be used safely. In fact, as I've said a few times now, all of these terrible weaknesses cannot be separated from deep learning's

strengths. And that strength, in fact, is a superpower that has not existed in any of our technology until this point.

CHAPTER SIX

KNOW-NOTHING MACHINES

LLMs know everything. That's why one enthusiast friend described hosting one at home as "archiving the internet." As I've also said, they literally know nothing. Isn't that a contradiction?

Except, of course, it isn't. You *can* use LLMs as an "archive." It is also true that LLMs do not explicitly store any information. When they do, it's actually considered a failure in the model, like when the Berkeley Artificial Intelligence Research (BAIR) found that GPT-2 could regurgitate certain people's phone numbers, work addresses, and emails.[1] This is partly for legal reasons: LLM companies want to avoid their models reproducing copyrighted content.

This concern isn't theoretical. The *New York Times* sued OpenAI for ChatGPT outputting certain articles verbatim.[2] But the other reason for this failure is that deep learning models like LLMs are meant to generalize, and they likely won't if they simply memorize and retrieve large amounts of data.

This speaks to how I described generative AI. It learns a statistical distribution, which can be as complex as language, and generates data from it that looks like it could come from the original distribution. That's how some use LLMs as a know-it-all advisor, when in truth it is a know-nothing machine—by design.

So are those who use LLMs as a "super search engine" making a serious error when using them for decisions that matter? It depends. Know-nothing machines are still useful when you can add a component that does know something—an oracle.

ALL YOU NEED IS AN ORACLE

In cryptography, oracles reveal information to an attacker. That information may be vague or incomplete—like pronouncements of the Greek Oracles of Delphi, from where the term is derived. Any information can vastly simplify a hacker's job.[3]

How? Imagine a password prompt that tells you if your wrong password was too long or too short. That seems like a useful feature for a forgetful user. Right? It's also a critical vulnerability that will make a password relatively easy to crack. With that bit of knowledge from that exposed password oracle, a hacker can rule out a vast number of password combinations. Specifically, it rules out all passwords of a different length. How many is that? It's not billions. It's not trillions. It's not even quadrillions. With a standard eight- to sixty-four-character password, by my calculations, the eliminated possibilities are a number that has one hundred and twenty-six zeros. That's many, many times the total number of atoms in the universe.[4] With a fairly average twelve-character password and ninety-five standard printable characters, that is still $95^{12} \approx 5 \times 10^{23}$ possibilities left. That's not a small number. But compared to 10^{126}? That's more than a 99.9999 percent reduction in password possibilities.

I have borrowed this concept from cryptography to understand when it's reasonable to use LLMs—or other deep learning systems—and when it's not. I feel it evocatively conveys the power of even incomplete verification. Even tiny amounts of information can make a universe's worth of difference. I mentioned before the cases of a relationship pros and cons list, a multi-million-dollar transaction, and calculations critical for estimating whether my wife or I could have a second child. I also mentioned that these were reasonable uses of LLMs.

To understand this, imagine writing an industry report as an analyst. You can use a verbose LLM product like OpenAI's Deep Research that goes off and independently writes full reports.[5] How useful would that be? As Ben Thompson from Stratechery found when he used it to analyze a friend's industry, it's quite useful—as an expert. Deep Research found a lot of basic information but completely missed a major player in the supply chain. That

entity happens to be a private company but is significant enough that any report that misses them would get a failing grade.[6]

He was able to catch this easily. He is, after all, an expert. If it was a real work product, he would have been able to take the broad report, edit it with the relevant information, and then have saved a significant amount of time. What if a junior analyst used it instead? That report likely would have gone out the door and been an embarrassment. The difference is Thompson can act as an oracle, and the junior analyst cannot.

The exact same thing is true with programmers. I've seen—and personally experienced—seasoned programmers *fly* with LLM help. I've also seen junior programmers quickly get crushed by bugs and errors introduced by the LLM—issues the junior didn't catch when reviewing the code.

We've talked about why lawyers, with a highly specific and high-stakes application, are particularly prone to ending up victims of hallucinations. I spoke with Devansh, cofounder of Iqidis (a legal tech start-up) and writer of a popular newsletter called AI Made Simple. He said the real trick in unlocking lawyer productivity is *not* to eliminate hallucinations—which is fundamentally impossible, as we've discussed. He shared that even aside from well-publicized cases of fake court precedents, LLMs can also fail more subtly. They can cite real cases but analyze them wrong. Instead, we should aim to make LLM-created legal content *easily auditable*—instead of a black box that necessitates rechecking everything. This enables a human expert, the lawyer, to quickly verify the work. That's also a core part of how Iqidis approaches the problem.[7]

My friend, who was making pros and cons for her relationship, already had a sense of what she wanted to do and obviously could have easily corrected something if the LLM spit out something nonsensical. The person who generated the report on his company's acquisition value obviously knew his own business. And besides, he sent it to his investors, including me, who looked it over. As for my wife and me, we were out and about with an energetic toddler, so having Deep Research write out the math for our IVF process and genetic condition was nice. I also could easily glance over the numbers—which I was going to do no matter what, given my distrust of

LLMs doing math—and recognize if something was off by a significant amount.

In all of these cases, the oracle makes it so the AI is useful. And in all of these cases, the oracle is human. It doesn't necessarily have to be, but with the limitations of deep learning and LLMs, it often will be. It's poetic. The combination that works is the human oracle and the know-nothing machine.

But the reason why it's useful to the human at all is it can do something our past tools, whether computerized or not, could never do—handle variation. The exact circumstances of the relationship, acquisition, or IVF are likely novel. The general outlines of these circumstances are not. And that's exactly where deep learning excels.

DEEP LEARNING'S SUPERPOWER

The history of automation has been standardization. Henry Ford famously quipped that customers could buy any color of Model T, as long as it's black.[8] The Green Revolution fed the world by creating vast monocultures of single crops.[9] Even in electronics, the adoption of standards like USB pushes uniformity for efficiency.[10] On a broader scale, manufacturing, factory farming, and standardized shipping containers—in other words, goods, food, and moving them around—are the building blocks of our modern society. And all of them have the express goal of driving out as much variation as possible. Up to this point in our history, the way to make things more efficient and scalable was by making them as uniform as possible.

Anything that lacked uniformity became a laggard in both technology adoption and productivity. Notably, as I've seen during the course of my investment work, industries like construction or food services suffer from the inability to adopt most productivity-enhancing technologies. No room is ever truly "square" (ninety degrees in each corner), and no lumber is ever that close to its specification. And even quick-service restaurants often have to accommodate exceptions ("hold the pickles" or "I'd like a double bacon cheeseburger without meat") that break attempts to roboticize with the technology of the twentieth century.

These industries have gained reputations for being stubborn and suspicious of technology. In my experience, that's because most gizmos that tech companies have pitched to them have been largely useless. The lack of strict, perfectly controlled environments means that we have to use the only means we have of handling variation, albeit small, which is to throw humans at the problem.

Then, deep learning arrived. Thus far, I've devoted far more space to the limitations of modern AI than its capabilities. That is partly a function of the moment we're in during this book's writing. There is boundless optimism—and fear—about what it can do. I've taken pains to emphasize that the very structure of what it is creates some significant limits. That being said, something that can dynamically handle actual variation as we've discussed, even if bounded, is a truly unique capability in the history of our technology. I mean *true* variation in general circumstances—not just tightly controlled variation like manufacturing tolerances (where even the word conveys how even that kind of variation is merely *tolerated*).

MACHINE INTUITION

My go-to description for lay audiences for how AI worked was always "machine intuition." I doubt I was the first to come up with it. I've encountered other experts in AI over the years who have independently described AI and machine learning that way as well. It's also a fairly obvious way to understand it if you've peered under the hood of these models.

"Machine intuition" predated "glorified autocomplete" for me. I mostly used it before ChatGPT completely changed the nontechnical public's conception of AI. No one ever asked about LLMs before ChatGPT, and since then, they're typically all anyone asks about. While "glorified autocomplete" is a more specific way to understand an LLM's true underlying mechanics, "machine intuition" still applies. It also applies to machine learning generally, not just the subset of deep learning.

I think the analogy is as perfect a fit as you can get because "machine intuition" captures the inherent lack of carefully considered logic behind these algorithms. They tend to draw conclusions accurately when the data

aligns with their prior experiences. But the beauty of intuition—machine or human—is that it isn't rigid.

You can understand that certain objects thrown at your head hurt, or not, depending on if it's a lead fishing weight or a foam ball. When you picture that in your mind, you aren't weighing the variables systematically: "The fishing weight is smaller but has higher density and thus higher overall weight. But in addition to that weight, it does not deform…" I highly doubt you're doing those physics calculations in your head for this. You just *know* that the fishing weight will hurt more. And unless you have been hit simultaneously by a fishing weight and a foam ball before, you also aren't recalling a specific memory. Nor are you likely recalling one instance of getting hit by a fishing weight, a separate instance of getting hit by a foam ball, and then comparing the two in your mind. None of that complicated calculation or memory retrieval work is necessary. You just *know*.

Similarly, machine learning algorithms—which tend to be all marketed as AI now—do not have a logical derivation for their answers or have traceable "memories" (data) that explain any single conclusion. It just *is*. Which means, like human intuition, it's a black box, with all of the challenges that come from it.

This fuzzy, squishy nature seems much more *human* than the typical cold precision we associate with machines. Just like how I've emphasized the limits of AI, I've also stressed how *not human* our current AI is. That being said, this *narrow aspect* does emulate an aspect of how we humans function, which is why it has the power to massively augment—or, in some specific cases, replace—humans to begin with.

Because, you see, humans can reason, use logic, feel, adapt, and do all sorts of things AI cannot. But in our basest modes, our minds operate on pattern matching.[11] I already described how my infant daughter could easily start recognizing "doggie" around one year of age. We know how to navigate unfamiliar houses because almost all of them have bedrooms and bathrooms in some layout. We can often "feel" that a place is dangerous or not, even if we can't quite put a finger on why. One theory in cognitive psychology of how this works is our minds form prototypes. We compare

what we experience with a stored “prototype” that is an amalgamation of our prior experiences.[12]

Pattern matching and prototypes are not how we became the dominant species on Earth. It isn’t how we discovered that the Earth revolves around the Sun. It isn’t how we put a man on the moon. It isn’t how we built our society as we know it. It isn’t even that different from what most animals do.

That being said, it isn’t something our tools have done before. While all of our past attempts at AI, back to Rosenblatt’s perceptron, were designed as learning machines, most computer systems did not practically operate this way on a large scale. And while technology companies like Google, as described to me by an early employee, used machine learning algorithms in their commercial products as early as the mid-2000s, what they could do does not look like what we can do now.

That intuition used to be much more limited. Machine intuition did not intuit images at a human level in the ImageNet challenge until Microsoft Research’s over-hundred-layer model in 2015.[13] Machine intuition did not (non-terribly) replace human translation jobs until the past few years—with the inflection point being 2016, when Google’s Neural Machine Translation System hit near-human performance.[14]

What changed? Deep learning, of course. But why? Machine intuition can describe most of machine learning. Just that step wasn’t enough. What else was needed? What did the “depth” in deep learning bring that changed so much and brought it closer to human pattern matching and prototypes?

It goes back to the title of this chapter. We’ve talked so much about how different modern AI is from expert systems, especially in its lack of true knowledge. But know-nothing is even more fundamental. The source of deep learning’s amazing power stems from its depth of layers giving it a quintessentially human characteristic.

Forgetting.

LEARNING BY FORGETTING

Recall what ultimately doomed expert systems of the 1980s and 1990s in practical use. The proximate cause of death was their inflexibility. But also remember how Goel described it—scale. It was too hard to scale systems that required careful curation of data and rules that guaranteed correct output. In theory, if you could have all the time and human labor in the world to do so, you *could* get systems as powerful as modern deep learning. In practice, that obviously negates many of the productivity benefits of AI.

Classical machine learning algorithms—meaning machine learning models that aren't deep learning—don't have that same excuse. As contemporaries, classical machine learning and deep learning both have access to the same levels of compute. Both have the same lineage, where one can pump massive amounts of post-internet data into them with relatively little fuss and limited human intermediation. Thus, compute and data resources are similar. The real difference must therefore lie in the models themselves—or, more precisely, in how they handle data.

You see, it isn't quite true that all machine learning algorithms can handle mass quantities of data well. Unlike expert systems, one *can* indiscriminately dump in data without human expert knowledge, but that doesn't guarantee a good outcome. Traditionally, machine learning models share a specific tradeoff with statistical models. This should not be surprising because they *are* statistical models. The tradeoff is so fundamental that every statistics student in college has it drilled into them. It certainly was for me, which means I can rattle it off in my sleep. So, here goes.

The tradeoff is between *bias* and *variance*. If you make a model too simple, it will have high bias and always be quite inaccurate—something no quantity of data can fix. More quality data won't hurt, but it won't help either. The model is too simplistic to benefit. The complexity and nuances of the problem are greater than that of the model itself. One might think that one could just make a model highly complex and solve this problem. Unfortunately, while that does solve bias, it introduces high *variance*. Specifically, that means the model is so complex that it can perfectly "memorize" the data it is trained on—also known as "overfitting." While the model will perform excellently on the training set (in fact, with enough

complexity, it can approach perfection), it will fail horribly on any new data. Rote memorization makes generalization, and hence real-world performance, of a model poor.

From this perspective, most deep learning models are *ridiculously* complex. Deep models have so many layers and parameters it's hard to believe they wouldn't simply store everything. This is part of why these huge, deep models take so much data to become useful to begin with. There are *so many parameters* that it takes a colossal amount of data for these models to start to perform. The head-scratcher is why these models, once trained, don't simply end up falling into the overfitting trap. Yet, empirically, they don't.

Deep learning models seem to bend the traditional rules of the bias-variance tradeoff. They are remarkably resilient to overfitting and incredibly amenable to *scaling*. The reason is the one we cited previously—forgetting.

The problem of overfitting lies in *memorization*. How I like to describe it is that the layers of a neural network act as "lossy compression." Audiophiles or photography nerds would know the concept. WAV files are compressed, with irreversible loss of fidelity, to MP3 files. RAW or PNG files compressed into JPEGs, with noticeable quality loss. In all of these cases, a high-fidelity master loses quality to be much smaller and easier to store but is still "good enough." The data within a neural network, passed from layer to layer, undergoes a similar compression. The details are lost, but the essence is kept. Again, this process is similar to the theory of how human pattern recognition works on prototypes.

Deep neural networks forget, but in the right way. They compress, like humans. When forgetting happens in the *wrong way*, performance collapses in a phenomenon called "catastrophic forgetting." When too much heterogeneous data is trained sequentially, early data can be "overwritten" instead of generalized—especially when training is poorly set up.[15]

The trick is neither rote memorizing data nor totally forgetting early data, but instead incorporating it all into a *generalizable prototype*.

We don’t *exactly* know how this works. Neural networks tend to be black boxes to human researchers viewing them from the outside, but we can theorize. One compelling idea, though still debated, is the Information Bottleneck Theory, which outlines what we just described. Because individual layers lack the complexity to fully memorize data, despite the full network having more than enough, neural networks are “forced” to only retain the “essence” of the data. Working together, multiple layers capture different fundamental concepts, which we cannot easily discern from the outside. They also have remarkable results that are resilient to overfitting (even if not totally immune) with vast datasets that would traditionally challenge classical models.[16]

Which, again, brings us back to why deep learning took off. *It scales*. Not just because of the compute. Not just because of the data. The models’ own ability to scale was a critical and necessary breakthrough.

WHAT TO REMEMBER ABOUT DEEP LEARNING

This finally gives us the essential pieces needed to understand why deep learning is special.

We walked through its primitive birth with the perceptron and the waxing and waning cycles against symbolic and more knowledge-based AI. We explored how computational power became critical with deep networks demanding heavy computation—especially through backpropagation—just to make training possible. We saw the cleverness involved in model development with transformers to parallelize calculations. Together, these advances made it possible for LLMs to scale enough to ingest, with only modest exaggeration, the entire internet. And finally, we delved into the fundamental reason these deep networks are so special. They, like us, forget the unimportant details and retain core abstractions—even in the face of unimaginable amounts of data. They *scale*.

Which, finally, brings us to machine intuition—something that still cannot truly innovate and adapt beyond its training set. But it *can* operate within prototypes it has seen and has adapted *within* the bounds of what it’s seen at human levels in vision, translation, conversation, and (as we’ll see) more.

It’s not human. But—despite my admonishments about anthropomorphism—it *can* do more human things than anything else before.

Now, some of the magic of LLMs *are* cheap tricks. Cracks like Rs in “strawberry” are trained out. ChatGPT o3 achieved an impressive overall score on ARC-AGI but did remarkably poorly on some easy abstract reasoning problems in the test. Discussions about LLM mastery of math proofs may largely be illusory. A clever paper tested all major models against the 2025 US Math Olympiad literal hours after it finished. The goal was to see how well these models would do without “data leakage,” where their training set inadvertently captured the answers to the test. It found that no models scored more than five percent. The models had essentially gotten the answer key from their training set. Even worse, as Gary Marcus has commented, all the models were totally confident they had done the problems right.[17] They were probably as confident as Claude was that there are two Rs in “strawberry.”

I say all of this not to diminish what LLMs *can* do. I say this to emphasize that LLMs are not magical and have distinct limitations that cause them to fall short of the hype I’ve seen in the media.

To overcome those limitations, we need to do more than simply throw more data into deep learning. At least in my corner of the AI world, we repeatedly ask whether we have hit “limits” to scaling. Should we rely solely on larger, more computationally intensive networks and increasing amounts of data? Maybe. But it’s much more likely we’ll find better and more innovative models, which will help us scale *alongside* more data and compute, instead of mere brute force reliance on ever greater volumes of data and compute.

BEYOND DEEP LEARNING

When we concluded discussing the historical back-and-forth between connectionism and symbolic AI, we alluded to synthesis, including systems that already exist like AlphaGo. Notably, AlphaProof and AlphaGeometry—also from Google DeepMind—achieved silver medal performances on the 2024 US Math Olympiad. Their performance, unlike that of LLMs, cannot be explained away by data leakage. It’s not a coincidence that

AlphaProof and AlphaGeometry are hybrid systems that have deep learning and symbolic provers.[18]

But we are also scaling LLMs *themselves* using hybrid techniques.

Reinforcement learning (RL) is a branch of machine learning that is one of the few directly geared toward "novel learning." It trains agents to make decisions using trial-and-error by trying certain actions and receiving rewards or punishments. Using this feedback, RL algorithms find paths to achieve their goals.[19] One of my favorite examples is a YouTube video by Peter Whidden where he teaches an RL agent to play Pokémon Red from scratch. You see the agent go from doing completely random actions, barely able to get out of the bedroom that the player starts in, to exploring and fighting Pokémon battles. It does not finish the game, but still performs impressively for an agent that literally knows nothing and learned everything through pure trial-and-error.[20]

That power isn't being used primarily to let LLMs wander outside of their space, but it is a key part of the "post-training" (tuning after running data through the model) pipeline that has made them so impressive. Post-training tunes out obvious problems in their responses. Nathan Lambert, a researcher on the topic at the Allen Institute on AI, described in his newsletter Interconnects the recipe of modern post-training as follows: Use extensive human preference data, scale "RLHF," use synthetic data, and curate high-quality data.[21]

"RLHF" is reinforcement learning from human feedback—the idea that we can combine RL techniques with human feedback to optimize language models. Human preferences can inform a reward model that then can help train LLMs to have output that is more "pleasing" to humans it interacts with.[22] Combined with synthetic data, helping further "pad" out the model and being somewhat selective about the data used completes the picture of Lambert's "recipe" for our purposes. The key here is that by using this technique that is outside deep learning, we are improving the results from what we'd achieve with deep learning alone.

It also mixes up traditional “supervised” (human labeled, as RLHF data is) and “unsupervised” (non-labeled, with the model automatically finding structure in its data, which LLMs were classified as early on) learning, which used to be one of the first distinctions between algorithms taught to new students.[23] Today, LLMs are more commonly considered “self-supervised” instead of unsupervised, though the point stands that few hard distinctions remain.[24]

As for the point Morffis made in the last chapter about LLMs getting added capabilities potentially overcoming their challenges, we *are* seeing ChatGPT and other LLMs doing so. I regularly see ChatGPT, Claude, and Llama run code, pull up background calculators, search the web, and more during my own sessions. Undoubtedly, these models will only get access to more tools as we progress. However, Morffis notes that models calling for tools are still imperfect if they hallucinate or have some randomness. In other words, even if the LLM is just a button-pusher, it’s a potentially unreliable one.

Even the “knowledge” part is getting fuzzier. Retrieval-augmented generation, known as RAG, provides an LLM with access to vetted data during the query.[25] Under the hood, it provides relevant information “just-in-time” like a search function. This approach *also* does not solve the LLM’s knowledge problems, but it can alleviate them and at least “push” them in the right direction.

Nevertheless, this *is* likely what we’ll see more and more with fewer distinctions or rigid dogmas between schools of AI. As one of my posts called “Scaling Is a Choice” pointed out in late 2024 during the height of press hysteria over “is AI scaling over” (the answer was no, by the way), scaling will likely require various cobbled-together techniques.[26] Even Moore’s Law—named after the cofounder of Intel, Gordon Moore—which observed in 1965 that transistors on a chip (and hence their speed) doubled roughly every two years, was not some natural phenomenon.[27] It required hard work with herculean engineering efforts to come up with new ways to scale chips every time we hit barriers—similar to what we’re seeing in AI now.[28]

We started in deep learning, which unlocked these capabilities, but you shouldn’t expect that we end here. The *core* strengths and weaknesses are still those of deep learning. And as we’ll see as we get deeper into applications, that strength can be quite considerable.

PART 2

AI IN THE WILD

CHAPTER SEVEN

AI AS A BUSINESS MODEL

As a venture investor, my specialty area is “deep tech.” Depending on who you ask, that could either mean everything in technology, especially now that AI is on the scene, or next to nothing. For my firm, “deep tech” refers to technology that is at the forefront of science, which typically possesses a substantial amount of intellectual property, and has the potential to completely remake the world, provided it scales up. You might assume that this means I spend my days constantly evaluating technology, digging through data reports, patents, and academic papers. You’d be (mostly) wrong. Make no mistake, I do spend time on those things, and the technology matters, but mostly as a checkbox. Has it been proven out enough or not? Subject-matter expertise and technical knowledge are required to answer that question, but once it is, that’s it. What really matters is the *business*.

One of the most overused aphorisms in Silicon Valley is, “First time founders care about product. Second time founders care about distribution.” It’s been attributed to many people. At minimum, Justin Kan—founder of Justin.tv, which later became Twitch—said it out loud on Twitter, which later became X.[1] I’ve parroted it myself from time to time.

Cliché or not, the reason it’s used so much is that it’s true. It’s easy to get enamored with the flash of a product—a beautiful interface, juicy animations, or some fancy technical breakthrough. Silicon Valley lumps this all under “product”—an umbrella term that includes design, technology, and how something is packaged up for customers—which confused me to no end when I first got here. Distribution refers to distribution channels and how you sell a product. However, in the spirit of the aphorism,

“distribution” can be seen as a stand-in for all of the less sexy parts, including business model, go-to-market, and defensibility.

Right now, much of the conversation around AI is blinded by technical flash. Far less attention is given to a more basic, pressing question: How will any of this actually make money?

Even if you aren’t an investor, this question matters. The aspects of this AI boom that will persist and most shape everyone’s lives aren’t necessarily the most theoretically powerful or technologically dazzling. They’re the ones that are the most economically viable with profitable business scale. They sustain themselves, and in a virtuous cycle of profits and reinvestment, they grow. Unprofitable ones rely on sponsors—big companies, government, or simply temporary hype—and collapse when that support is withdrawn. It’s like pushing a boulder uphill. When that constant effort is withdrawn, it rolls back down, potentially wiping out whoever is in its path.

In 2025, Meta announced up to seventy-two billion dollars in capital expenditures, primarily on AI.[2] Along those lines, Goldman Sachs has projected that AI will consume over one trillion dollars in the coming years. Yet it has yielded (relative to investments) minimal results thus far.[3] While OpenAI is now generating revenue—a projected $12.7 billion in 2025—it’s still far from profitable.[4]

Even bigger bets are on the horizon. OpenAI’s project Stargate, a mega-scale compute facility backed by SoftBank and the UAE, announced when President Trump came into office in 2025, is just one example.[5] Meta and Microsoft are racing to secure nuclear power to support their AI data centers.[6] These efforts only make sense if they eventually generate a return on investment, making more money than they cost. So far, that’s far from reality and certainly not guaranteed.

Is this all worth it? Once again, because it matters: How will any of this make money?

To understand that question, we need to actually understand several concepts underlying the structure of AI as a technology. To start, we will

revisit a now-familiar one.

MODELS, COMPUTE, AND DATA

Although I've drawn a sharp distinction between the "product" and the business side of technology companies, the nature of the product inevitably shapes the business itself. Selling Microsoft's suite of Office products is obviously different from selling homemade lemonade. How much a product can be sold for, how it can be sold (downloaded, online sales, or brick-and-mortar retail), and how easy it is to knock off are all factors inherent to a specific product and influence how a business can be built around it.

One of the challenges for both entrepreneurs and investors in start-ups—like venture capitalists—is that the list of what matters can be virtually endless. Time and focus, unfortunately, are finite. Given that, it's important to sort out what matters most and what matters less.

Of course, new technology can also enable business models that were previously unviable. Webvan is a classic example of a "dumb idea" from the DotCom era. That idea was to deliver groceries to homes via the internet. It was worth ten billion dollars in November 1999 and shuttered its doors in 2001 after burning through more than one billion dollars but still falling short in customers and revenue.[7] With essentially the same vision, Instacart went public in September 2023 with an estimated value of (in a cosmic parallel) ten billion dollars.[8] Years later, with smartphones and faster mobile networks—essentially, a "better," ubiquitous version of the internet—a dumb idea became a brilliant one. Understanding the nature of a technology provides an understanding of what it can enable as well as its competitive profile.

For AI, the core framework I've already used multiple times in this book is models, compute, and data. Using that framework is the best way to quickly understand what can or can't be a successful business within AI.

Models are the architecture, algorithms, and techniques used within the AI field. Although it's sometimes used in a narrow sense—referring to only the literal architecture of a machine learning algorithm—I use the term broadly to encompass all generalizable technical knowledge in the field. I'm

certainly not alone in doing so, given how blurred the lines often are between what is technically part of the formal architecture or not—for example, post-training techniques applied to LLMs. I'd prefer to think of it as the complete "recipe" that is used to create a machine learning model.

Compute is the new, hip shorthand word for computing power. Previously a rarely used verb, it's now a noun used everywhere—from mainstream newspapers to government documents and press releases from tech companies. Its rise as an industry buzzword is a sign of just how central compute has become. Most compute, as we will see in the next chapter, comes in the form of semiconductor-based GPUs, but I will broadly use the term to refer to the physical systems that AI runs on. To extend the cooking analogy, it's the kitchen appliances, pans, and utensils used to make food.

Data is the application-specific information fed into models and run on compute—even if the "application" is as broad as human language, in the case of LLMs. Many AI algorithms, including deep neural networks, are a completely blank slate. Data gives form to that slate, like a die stamping a flat piece of sheet metal, leaving behind its impression to be reused, refined, and replicated. To finalize our cooking analogy, it is our raw ingredients.

But if everyone has models, compute, and data, what actually sets an AI company apart?

Models are typically shared openly and improve rapidly due to global collaboration. As we've seen numerous times, Google has published many state-of-the-art models in AI.

Compute is expensive but still accessible. Anyone with enough capital can buy GPUs. Elon Musk's xAI recently joining the frontier model club is illustrative of this point. The company certainly has strong talent, but the way they caught up is essentially by raising six billion dollars to buy one hundred thousand Nvidia GPUs.[9] That's more than pocket change, but remember that it's a mere fraction of the seventy-two billion dollars Meta is projecting to spend *alone* in 2025. While you or I couldn't go out and buy that many GPUs, it's certainly within reach of most Fortune 100 companies.

That leaves data—the final pillar where AI companies can differentiate themselves.

EATING BIG DATA

The term “big data” originated in 1999 with the rise of the internet as we know it today.[10] Companies like Google and OpenAI have ingested this data from the publicly reachable web to train their AI. Meta, at least (but it is unlikely alone), has even gobbled up troves of pirated books from online databases like LibGen.[11] Blogs, social media, and news sites are bountiful seas of text. Platforms like Instagram and YouTube provide an infinite scroll of photos and videos. Podcasts and audiobooks likely provide millennia of spoken voice. It’s fair to say, given the incredible volume of data required to train LLMs of the size used today, that we couldn’t have done it before the existence of such easily harvested—or stolen—data.

But this data, scraped from the public square of the internet, is largely the same for everyone. That means all these LLM model companies have similar models, compute, and data, which also means all of these LLMs should be broadly interchangeable. Users might prefer slightly different tones, verbosity levels, or interfaces, but most frontier LLMs struggle to differentiate themselves and have faced diminishing returns in achieving greater real-world performance noticeable to average users.[12]

Of course, many casual users of AI have never used anything except ChatGPT, at least from my conversations with friends, family, and random people I meet. There is power in brand recognition, after all. If there wasn’t, OpenAI wouldn’t have spent fourteen million dollars on a Super Bowl ad in February of 2025.[13] However, OpenAI is still unprofitable, not just as a company, but even on a *per-user* basis for their pro-tier two-hundred-dollar subscription.[14] Effectively, they’re subsidizing their product. What happens if they *really* start charging enough to make money, especially when they have competitors like Meta that are giving away their competitive product for free, as of this writing?

Brand has value. You’re willing to pay more for a Coca-Cola than a store brand. But *how* much more you’re willing to pay has limits. A little bit more? Sure. Twice as much? Maybe. Ten times? You might start

reconsidering. Competitors with similar products for lower prices or free exist, which puts a ceiling on how much a company like OpenAI can charge for ChatGPT.

Even if a company had some unique data, the sheer scale of the open internet renders most proprietary datasets marginal—at least for text, audio, images, or video. Even "multi-modal" models (those with capabilities in images/audio) like the latest versions of ChatGPT are unlikely to differentiate themselves with just "more data."[15] But LLMs are not the only form of AI.

DEFENSIBILITY IN DATA—FRICTION IS GOOD

The bloodstream of the internet is general text, audio, images, and video. It's so ubiquitous that it is difficult for any one company to get an "edge" in it, but many types of data are entirely inaccessible to the public. Sticking with the web for a moment, Amazon has a lot of proprietary data on customers and buying habits within its platform. That being said, between other large e-commerce platforms, third-party cookies, "free" financial apps that ask for bank/credit card statement access, and coupon plugins, that data might not be that locked-down or unique.

But there's a whole world of data *not* on the internet. Much of that data is locked up in proprietary corporate servers. Significant amounts of logistics and warehouse data are housed only within a single company's systems. Medical data, although technically located on electronic medical records (EMRs) in the US, infamously do not "play nice" between systems and end up effectively locked up anyway.[16] Even if it wasn't, data errors and contradictions make it difficult to make heads or tails out of it without a lot of specific context.[17] Finally, data in biotech, whether in genomics or biomanufacturing, key to unlocking the blueprints of life, have public repositories largely populated with junk—a repeated complaint I heard when I spoke at a workshop at the National Academies in Washington, DC, on AI and biotechnology.[18]

Whether they suffer from errors or are confined to inaccessible systems, most of these cases touch the physical world in some sort of way, shape, or

form. As the popular saying goes, garbage in, garbage out. In this case, it's sometimes nothing in, nothing out.

Digital data is constantly generated. It is the virtual footprints of billions clicking, reading, swiping, and scrolling. Whether it's e-commerce, search, social media, dating apps, or blogs, the digital world is hermetically sealed, automatically generating and capturing data with no friction. Data touching the physical world does not behave in this way. Someone typically has to record the data, making the translation from physical to digital. Even with automated record-keeping, when it exists, no sensor or measurement is ever perfect. This friction is painful—limiting how much data is generated, undermining its reliability, and raising the cost of generating your own. After all, the ultimate example of needing to gather highly reliable data in the real world is pharmaceutical drug trials, and those can cost nearly one billion dollars per drug candidate.[19]

This friction can obviously be a curse. No TV, radio, or billboard advertiser —despite being huge in their Mad Men heyday—could compare with the level of dominance of the likes of Google and Facebook/Meta. These digital native companies had the ability to cheaply and precisely track and target user purchasing behavior. Their traditional counterparts needed to somehow observe impact after the fact or by surveys in the physical world, with all the messiness and imprecision that entails.

That's one thing the internet revolution shares with the AI one. The internet allowed for mass customization. Niche interests had blogs and influencers. Niche brands could make direct-to-consumer web stores. Giant companies like Google and Facebook, with their massive but precise ad networks, could bring them all together. AI's superpower, if you remember, is the ability to automatically handle variation, which even for the internet was tackled by humans sorting themselves with their behavior.

That's why friction can—and is—a blessing for AI companies. Oncoustics, a company I backed through my firm, is an example. You may be familiar from either personal experience or movies with ultrasounds, usually associated with pregnant women and fuzzy images of a fetus on a screen. However, that image is derived from *sound* data (hence, "ultrasound").

From my discussions with clinicians while researching the company, almost no hospitals ever keep the raw signals. Few even retain the live feed. Most keep only a few specifically captured snapshots.

However, the *raw data* has an incredible amount of information that cannot be seen on screen. In particular, one can use AI to derive the flexibility of an organ like the liver. Clinicians can then use that information to understand liver health and diagnose conditions like fatty liver or fibrosis. How do you pick up the flexibility of an organ with a fuzzy, 2D black-and-white image? You can't. When it is turned into an image, that data is lost. But no human is going to be looking at raw ultrasound waves. Normally, it takes specialized, purpose-built devices to detect liver stiffness using ultrasound. With AI, Oncoustics can use the normally low-fidelity information derived from a portable ultrasound to reach the accuracy of that expensive specialized device.

The point is not to talk up the company but to show that this feat can be accomplished with today's AI. Anyone, in fact, could do it if they had the data, which they don't. The company has been gathering their dataset from partner institutions for years. Competitors can't simply buy data since raw data is almost never kept. The very expense and pain—the *friction*—of trying to set up processes to gather the same dataset is the essence of the company's defensibility. Is it impossible with enough time and money? No, of course not. It's simply unlikely to be economical. Everyone has the same models. Everyone has the same compute. But data? That's the castle, with a very real moat.

Businesses like this are inherently *quite specific*. Internet companies like Google had the benefit of being the focal points of online behavior for *anything*, which helped make them so incredibly valuable. Everything from online shopping to PhD research to data would walk through their digital doors. They were extremely *horizontal* in capturing multiple industries and segments. High-friction, specific data is extremely *vertical*. However, like Webvan to Instacart, many businesses were not viable before the advent of modern AI but now are. Even simple things that used to require humans—and hence, more often than not, just weren't done at all—can now be done economically with the support of AI.

Many of these simple things can be executed using LLMs. I can ask for long research reports on random industries, vague trading ideas, or just deep dives into specific topics that pop into my head. The internet almost certainly would have blogs or posts about all of these. That's where LLMs are at least partly sourcing their information from, after all. But they can now put them in an easily digestible format I'd otherwise need to (but never would) hire someone to do. I can't blindly trust the information, but that would be the case even if I hired a human gofer to research my every spur-of-the-moment shower thought.

I have my pick of the litter for which LLM to do this. But if I were to need to do highly specific disease screening, very specific advanced materials synthesis, or any number of *vertical applications*—which we'll absolutely see more of—I might have only one choice, which inherently commands pricing power, even if the total size of the market is smaller. These vertical moats are narrower in market size and breadth than the horizontal sprawl of web giants, but they also work well in this AI era where we can generate far more value out of data.

That does get to another point where AI has a distinct *disadvantage* relative to internet-era software companies: cost.

AI ISN'T "SOFTWARE"

Trivially speaking, AI is written using software. Software can even use a very similar framework as we've discussed for AI. You can write software—and even run simple statistical models within it—using compute and applying data to it. However, typically speaking, a canonical "software-as-a-service" (SaaS) company, which has dominated venture capital for years, has very low start-up costs and practically no marginal costs.

The translation of this is that two programmers in a garage can write code to make a product with relatively little money upfront. After that, each new customer costs the software company next to nothing. It's just software served online, which means no shipping, manufacturing, or any real cost at all other than the (often cloud) server. This can then scale to thousands, millions, or billions of customers and be a remarkably profitable business. Sure, as the company scales, you need more marketing spend and

headcount, but the costs don’t scale in line with revenue, which means profit margins can be greater than almost any other type of business. It’s clear why venture as an industry has loved this model so much.

AI is written with software. But it doesn’t look like this at all. It should already be obvious with the mind-bogglingly large amounts of spending I discussed at the beginning of the chapter.

Start-up costs certainly are not low. An AI company needs both data and *training*. Data can often be quite expensive. Putting aside high-friction data, even if all of the LLM model companies essentially stole all of their data from the internet, that would be a massive engineering and labor task of finding and building tools to scrape it. That’s not cheap. And that’s just for the data. Training, remember, is feeding the data in to correctly set the weights within the model, which is obviously necessary to make the AI system useful. OpenAI’s GPT-4 was estimated to cost seventy-eight million dollars to train. Google’s Gemini Ultra was estimated to cost $191 million.[20]

Training costs for frontier models have been going up over time and are expected to keep doing so. Some have used Chinese models as a counterexample. However, most erroneously cite the cost of DeepSeek’s frontier model, v3, as $5.3 million. To clarify for readers, that figure refers only to the last training run. It excludes all experimentation, hyperparameter tuning, and failed iterations that make up the bulk of real-world costs. If actually compared apples-to-apples, DeepSeek may be somewhat cheaper than other frontier models but is still in the same ballpark.[21] That means the point definitely stands.

However, it’s not just start-up costs. *Inference*, meaning running the AI, costs money as well. Although inference costs have come down, in 2023, SemiAnalysis estimated that ChatGPT’s inference cost was less than half a cent.[22] That may seem negligible. After all, it’s less than a cent. But SaaS companies often *charge* $0.001 or $0.002 for their equivalent services and have huge margins, which means their *costs* are a fraction of that amount.[23]

Also, just imagine millions of high school students typing random queries into ChatGPT every few minutes. It really doesn’t take long for that figure

to become huge, which helps explain, in part, why OpenAI has needed to raise so much money. The majority of this cost is actually power, which, by my calculations, is the energy cost of running an air conditioner for a few seconds—each query.[24] Again, it doesn't sound like much, but the equivalent SaaS query is practically immeasurable. This, again, is why Microsoft and Meta are looking to reactivate or build nuclear power plants for their AI data centers.[25]

AI is expensive to both start *and* run. We've talked about how incredible the capabilities of deep learning really are (even with its limitations), but with these kinds of costs, AI companies need to generate significant value—and command meaningful pricing power—to make the math work. High-friction, proprietary data creates that pricing power. But does this mean LLM model companies are doomed?

NATURAL MONOPOLY OR TOO CHEAP TO METER

Ironically, "foundational models" like ChatGPT and other LLMs can go two radically different but plausible directions. Both are supported by trends that we see.

On one hand, we see the AI all-in training costs per model have been skyrocketing. In 2017, the cost of the original transformer model was around a thousand dollars. In 2020, OpenAI's GPT-3 was $4.3 million. And, of course, as already pointed out, Gemini Ultra was nearly two hundred million dollars.[26] While it's gotten harder to specifically estimate what any specific model costs, it's clear from the billions being sunk into infrastructure that the spending curve hasn't stopped.

On the other hand, inference costs, thanks to more efficient hardware and model techniques, have been dropping. Andreessen Horowitz estimates that inference costs have dropped by a factor of a thousand between 2021 and 2024.[27] While AI inference costs are still higher than SaaS costs, we have been seeing them drop significantly year by year.

So are things getting more expensive or cheaper? Both. Hardware and techniques are both getting more refined. As such, inference costs *and* training costs have fallen. Ark Invest estimates that model training costs

have been and will continue to fall at a pace of 75 percent per year.[28] But doesn't that contradict what I said about model training? It doesn't. The *exact same* model that cost one hundred million dollars to train last year now will cost twenty-five million dollars to train next year. However, the new frontier models that are driving the cost *are not the same models*. Our capabilities are growing but so is our ambition and thus our costs.

Most readers are likely familiar with Moore's Law and the march of rapidly falling computing costs over our lifetimes. It almost seems like a true "law" of technology.

Fewer are likely familiar with exponentially increasing cost incurred to stay on the technical frontier. Ironically, the same industry that drove Moore's Law provides us the clearest example—semiconductor fabricators (or better known as "fabs"). In 1970, the average cost of a fab was thirty-one million dollars (2024 dollars). In 2010, that cost was around ten billion dollars (2024 dollars).[29]

As eye-popping as these figures are, the true "catch-up" cost of those who want to challenge the two main leading-edge fabs—TSMC (Taiwan Semiconductor Manufacturing Corporation) and Samsung—is even steeper.

Rafic Makki was an executive fellow at GlobalFoundries, and we'll hear more from him shortly. He explains that it isn't just the cost of a new fab. "The challenge of producing chips at the leading edge is you need decades of failure analysis data. Let's call it learning. You cannot get to the next node unless you're very good at the current node. And you cannot be good at the current node unless you were good at the previous node."

An aspirant to entering this rarified group can buy the equipment. They also can steal experienced people. But they also have to sink in the raw cost of unprofitably rolling down the cost curve and gaining experience and data, adding to the already ruinous cost of just the fabs alone.

Classically, natural monopolies are utilities. High fixed costs, relatively low marginal costs, and finite markets—usually dictated by geographic areas—lead to the outcome that a market is simply not large enough for competitors to profitably enter the market, so they don't. Hence a single

player (or at most a few) can charge monopoly prices.[30] Fabs, with extraordinarily high fixed costs, relatively low marginal costs, and only a few large customers, fit this definition. LLM model companies, if frontier model costs keep scaling as they have, could eventually *also* fall into this category. Or they could actually see their costs drop. Counterintuitively, that might actually hurt their ability to command monopoly profits.

To summarize, the two possibilities are:

1. Frontier models continue to rapidly scale in cost, leading to a natural monopoly.
2. Training costs stay flat or collapse, which ensures total commodification of models.

As said, the paths are radically different. Even if the model companies do become natural monopolies, it could be a pyrrhic victory. Monopolies tend to get regulated. Whether LLM model companies collapse into commodity providers or ascend into utilities depends on the slope of the cost curve.

SO WHAT ACTUALLY WINS?

In the heyday of Angelfire and Geocities sites, IRC chats, and MySpace/Xangas—in other words, my childhood—it would have been hard to imagine the wild, crazy, decentralized internet becoming largely controlled by a few giant companies. There were certainly signs that it could happen, and some did theorize (most notably, Ben Thompson with Aggregation Theory) that the nature of zero marginal costs would lead to mass consolidation. But the exact shape of today would have still been hard to predict.[31]

Similarly, it's difficult to know what will emerge from our current turbulent AI boom. It's easier, in a way, to predict what *can't* work rather than what can because structural limits can be foreseen better than visionary innovation. For example, as it stands, LLM model companies aren't likely going to make a return on the investment put into them because they don't have distinctive models, compute, or data. But who really knows?

Still, it's a bit too easy to say, "I don't know." And in truth, we have covered a few different ways of making money as an AI company.

- **Data Monopolies**: Companies that own high-friction data can build defensibility for their AI within specific verticals. These usually touch the physical world in some form.
- **Natural Monopolies from Costs**: In my opinion, this is the best-case scenario for many of the LLM model companies, which otherwise would not have strong differentiation. Eventually, if frontier models' cost curves look similar to semiconductor fabs, these companies can have pricing power through industry structure.
- **Traditional Product Companies**: Not every successful AI company needs to have a monopoly. Some companies may simply embody the DNA of recent Silicon Valley and create better AI-related products, featuring traditionally superior design, user experience, and interfaces. I've heard some derisively call these "GPT wrappers" (a thin layer around ChatGPT/LLMs), and it is true that if one company quickly builds a wrapper around an existing AI product, another can easily do the same. They don't benefit from structural protections, will need to compete harder just to generate the same level of returns, and will need to continually innovate to stay ahead. However, many great companies have succeeded by doing just that.

Of course, governments may intervene to curb or support monopolies for strategic reasons if they arise. The latter is quite possible given the importance of AI for national power, which we'll also cover soon. For now, we should dive deeper into what drives many of these dynamics and helped enable modern AI—its hardware.

CHAPTER EIGHT

AI HARDWARE AND PARALLELISM

It’s difficult to appreciate how complex modern computing hardware is. Rafic Makki—head technologist at Mubadala Capital, former executive fellow at GlobalFoundries, and former professor of electrical engineering—described making a perfect chip to me this way when we sat for an interview:

“Imagine you have a map of all the roads in the entire United States. Now, I make you sit down and draw the entire thing, but with every road to every little village. There’s a pothole somewhere, but you don’t know where it is. How do you find it? Okay, if that’s hard, how are you going to find a defect if you have one hundred billion transistors and trillions of interconnections?

“With the latest, leading-edge chips, features (elements on a chip) can be twenty nanometers or smaller. A human hair is around one hundred thousand nanometers in size. In modern chips, there are billions of transistors in less than a square centimeter, which all have to be connected. What’s the chance you get every wire right?”

We haven’t quite reached the level of pushing around atoms for semiconductors, but we’re really not that far off. I briefly described the incredible rise in the cost of semiconductor fabs last chapter, and this level of microscopic precision is the reason. Precision at such an incredibly tiny (physical) scale at such an incredibly massive (manufacturing) scale costs a lot.

Moore’s Law was an observation that transistor density—and hence, computing power—would double around every two years.[1] Over the course

of nearly half a century since 1980, this has held true. This computational rocket ship enabled our modern world that has computing embedded everywhere. This rocket ship slowing down and sputtering to a halt is how we ended up with the structure of internet companies and modern AI architecture we have now.

THE DEATH OF MOORE'S LAW

AMD founder Jerry Sanders has often been cited for his quip, "Real men have fabs."[2] That made sense when the cost was relatively affordable for a normal company. That's less reasonable when a single fab—like Samsung's new Taylor, Texas, facility slated to open in 2026—will cost nearly twenty-five billion dollars.[3] This skyrocketing cost resulted from the awe-inspiring complexity we described, ensuring that few companies in the world would even attempt to build fabs on the leading edge anymore.

The world benefited from exponential computing power growth, which also led to the collapse in its cost. In the 1980s, the cost of a gigaflop (one billion floating-point operations per second) was $18.7 million. In 2000, it was $640. In 2017, it was $0.03.[4] You may not viscerally feel it—after all, your laptop probably didn't cost a dollar—but that's just because, like frontier model developers, we simply found more reasons to use ever greater compute.

All good things come to an end, though, and even with exponentially rising fab costs, we can't maintain Moore's Law the way we once did. In fact, Moore's Law was already breaking down in the mid-2000s. Dennard scaling, the principle by which smaller transistors meant less power and heat, abruptly hit a wall due to physical size limits around 2005. The industry was reliant on that to run processors faster without burning a hole in themselves. Dennard scaling *was* what people thought of with Moore's Law.

The demise of Dennard scaling changed the picture significantly. Instead of smaller, faster processors, chips instead got *bigger* between 2005 and 2020. They became multicore, with more processors working in parallel as the means of improving performance. Finally, from 2020, we entered the world of three-dimensional chips, advanced materials, and less straightforward

innovations.[5] We're still rapidly growing in computing power, but the source of the power isn't just from computers simply getting faster anymore. We stopped getting smaller, faster chips, which, incidentally, meant we could put more of them in the same form factor. Moore's Law onward meant that we merely got *more* of them.

This radical shift meant that utilizing more chips in *parallel* became key because that was the only way to actually "speed up." Parallelism now pervades the entire computing stack, from the highest levels to the lowest.

PERILS OF PARALLELISM

Human beings don't intuit exponential scales well. Many stories tell of the invention of chess, but one involves a young man inventing the game to the delight of an Indian king. The king asked the man what reward he would have. Instead of gold, jewels, a palace, or even a province, the man asked for doubling grains of wheat for each square on the board. One grain on the first, two on the second, four on the third, and so on until the sixty-fourth. The king readily agreed, and the royal court laughed at the stupidity of the man, at least until the mathematicians calculated what was actually owed. The wheat owed would fill a cylinder with the diameter of a large city and be ten times the height of the Himalayas. All of the fields of India, sown with wheat, could not deliver on what was promised for two hundred thousand years.[6]

Moore's Law gave us that kind of heady doubling of pure computing speed for years. A software business aiming to make a competitive product, instead of spending time optimizing, was often better served simply writing something sloppier that hit the market faster. What did it matter? Next year's hardware would comfortably handle what barely ran today. As said, that held for a long time but stopped being as straightforward in the mid-2020s.

Why does that matter? Similar to the story about the grains of wheat, it's hard to know how incredibly fast our computing speed would be today if we were actually able to continue Dennard scaling. Faster chips lead to faster software speeds in a straightforward way. More computing power in parallel chips does not. Software needs to be written in very specific ways

to take advantage of parallel computing power. Between the complaints of friends who are professional developers and my own experience, it's not particularly easy to write that kind of code.

Aside from tooling issues and indecipherable bugs, the problems themselves simply can't be architected in the same fashion. The fundamental characteristics of parallelism prevent it.

To understand this better, imagine a very busy, urban burger shop that has to service a massive lunch rush from workers in neighboring office buildings. This burger shop has two ways to staff the shop. One option is hiring a superhero that has super speed. This single worker can assemble twenty burgers alone in a minute and simply does it one by one at superhuman speed. A second option is hiring a team of twenty, who can each assemble a single burger in a minute. For better or worse, the second option is *not* the same.

Now, let's be clear, if the problem were perfectly parallel (in computer science terms, *embarrassingly parallel*), as in each worker can be an assembly line of one, reaching for a bun in their own stack, then grilling a burger patty on their own little burner, and then squeezing sauce from their own individual set, then yes, it would be the same. In reality, the store probably does not have twenty individual burners, twenty individual sauce sets, and so on. Relative to a single superhuman, the twenty workers probably need to divide up work into individual stations—if nothing else, for space and capacity reasons at each one.

At theoretical maximum, those twenty workers can match the superhero. In reality, as anyone who has worked on a group project can attest, more people doesn't lead to a linear scaling of work output. Oftentimes, it's much more efficient just to do something yourself. Those twenty people will likely fall *far* short of the single superhero. This isn't just because everyone is lazy. There are *real* problems with dividing work.

The problems come down to communication, coordination, and various forms of waiting. No matter what, a team does not share brains or hands. Some level of time and resources has to be dedicated to talking or planning

and literally passing work product (buns, patties, and the like) around. Someone will always be waiting around at some point. Not everyone does everything at the same speed, and, especially with poor coordination, not everyone is literally doing the same amount of work. By definition, that means not all of the theoretical productivity of the twenty is being achieved. Finally, even putting aside these inefficiencies, the speed at which a burger gets made will never be less than the longest single bottleneck process in the work chain.

The same problems occur in parallel computing. Communication taxes the process—most significantly, passing around data, but sometimes literally just communicating status between processors. Without perfect parallel design, processes will tend to get imbalanced, with some processors churning at max output while others stay idle. Even with perfect design in splitting a problem, some calculations inevitably take longer (even if divided as best as one can beforehand), and other parts will need to wait for pieces to be done from prior processes. Finally, system performance in parallel computing can never exceed its most significant serial bottleneck. That makes intuitive sense. If you aren't moving faster, you cannot finish in a shorter time than the longest single task that must be executed. This principle is called Amdahl's Law.[7]

Just for fun, I'll give you another set of computer science terms. A superhero's speed provides *low latency* while a team offers *higher throughput*. The latter can achieve the same results as the former but only if parallelism is done perfectly.

While the dilemma of hiring a superhero or a team isn't a realistic one for a burger shop, it is one in computing. The most straightforward version of this is a CPU (central processing unit) versus a GPU (graphical processing unit).

GPUS ARE PARALLEL PROCESSORS

For many years, a personal computer's CPU determined whether or not it was "fast" or "slow." A CPU rapidly processes computations sequentially and, when chips were still subject to Dennard scaling and increasing their clock speeds, would literally execute instructions faster—like a speedster

superhero. GPUs really didn't matter for the typical user who mainly browsed the internet and read email. They mainly were useful for specialized professions like video or music editors, engineers who used computer-aided design (CAD), or gamers who wanted to play the most cutting-edge titles. In a way, that's *still* true today.

After all of the discussion as to the pitfalls of parallel processing, why are these GPUs useful at all, though? In large part, this comes down to certain problems being parallel anyway—where there is no (or at least less) dependency on prior computations for subsequent computations. And, importantly, we aren't dealing with a twenty-time speedster versus twenty bodies. Today's CPUs often are multicore, so there are actually multiple speedy heroes. AMD's Ryzen CPU series in 2025 had between six and sixteen cores.[8] Seems like a lot, right? Nvidia's consumer GeForce GPU series starts at *3,840* cores for their lowest-end RTX 5060 and goes up to *21,760* cores for the RTX 5090.[9] Like our analogy, each GPU core has a fraction of the performance and speed of a CPU core. However, there are *many more* GPU cores than CPU cores. Of course, that makes coordination and synchronization losses potentially even worse if problems aren't designed correctly to capitalize on parallelism, which, as said, is hard.

That's why it's helpful that with transformers and even the basic structure of how most neural networks are trained, parallelism dominates. But, of course, that's a chicken-and-egg question because, if you remember, one of AlexNet's big leaps forward, other than utilizing neural networks for computer vision, was using GPUs to accelerate it. Deep learning grew up *around* parallelism, and that's why it's important enough to cover it.

It's also partly why Nvidia has such a great business and has become one of the most valuable companies in the world since the AI boom started.

CUDA AND SCIENTIFIC COMPUTING

Nvidia launched CUDA in November of 2006 with their GeForce G80, which had 128 CUDA cores—GPU cores with extra hardware to support CUDA's functionality. What was CUDA? It stands for Compute Unified Device Architecture and is a general-purpose programming framework and

set of tools that make it much easier to build parallel applications on Nvidia chips.[10]

Software typically “feels” easy. As Marc Andreessen famously quipped, “It’s called hardware for a reason—it is hard.”[11] That’s true, but hardware is useless without software to run on it. Low-level software, the kind that maximizes hardware potential, is also quite hard. It’s also extraordinarily important, given it is the building blocks upon which all other software sits.

So how hard is this low-level software? It was so hard that when Nvidia pushed to release CUDA and convert all of their GPUs to be compatible, their gross margin (one measure of profitability) fell from 45.6 percent in 2008 to 35.4 percent in 2010. Combined with the global financial crisis, their stock fell by a catastrophic 80 percent between 2007 and 2008. It was a call from the very top by Jensen Huang, CEO and founder of Nvidia who acknowledged that the profitability impact of CUDA “started out poor and it got worse,” but he persevered because, “We were convinced that accelerated computing would solve problems that normal computers couldn’t. We had to make that sacrifice. I had a deep belief in its potential.”[12] With Nvidia’s dominance in AI and global markets, his faith obviously panned out in the long term.

CUDA’s original users were scientists, academics, and researchers—from chemical engineers, biologists, and pharmacologists in everything from life sciences to oil prospecting. They were willing to work through bugs in the software in coordination with Nvidia to refine CUDA for scientific computing. Before it rolled out AI in the 2020s, CUDA had been battle-tested for over a decade in real-world applications, which made gaining market traction for latecomers like AMD, Cerebras, and similar quite difficult. As one user stated, “There’s more than just putting your program onto different chips. It’s not simple. As a user, I can tell you they never work perfectly. It’s not worth it. I’m working on 99 problems for my start-up already. I don’t need another problem.”[13]

NVIDIA’S DOMINANCE

I’m usually quite hesitant to base too much of what I write on any specific technological stack, other than as a historical note. After all, the field of AI

is moving rapidly. Nvidia and CUDA are somewhat of an exception. Its dominance of the field—I've literally seen AI developers wrinkle their noses in disgust at the prospect of trying to use AMD GPUs—has impacted the stock market (which I care less about), national security (which we will discuss), and the literal landscape of AI (which matters most here).

A part of it has to do with what we're talking about—parallelism. I can speak from personal experience, having taken graduate-level courses in high-performance computing—and implemented highly parallel production systems myself—that alternatives to CUDA are a nightmare. Certain investors and developers have pushed back to me that platforms like PyTorch (an AI library supported by Meta) would obviate the need for CUDA. Others have mentioned that specific chips like Google's TPU (tensor processing unit) and the growing pile of AI hardware start-ups would break Nvidia's dominance. I would assert—and Nvidia's continued dominance would suggest—that they have been wrong.

AI is much more mainstream these days, and many more people have entered the field. I am primarily an investor and strategic voice in companies now. As such, it's a bit odd to say, but my perspective partly comes from being much earlier to the field than many now running around in it. I was never some kind of leading researcher—far from it; my days of being in PhD-level machine learning seminars at Berkeley came from me being "creative" with my MBA elective choices. I still remember being somewhat slack-jawed and facing the silent bafflement from others (actual PhDs and professors) in *multiple* seminar rooms when I was asked to discuss my dissertation topic. That being said, I cut my teeth and bled with some who were on the frontier at the time.

As such, I have a better sense of what's necessary on the leading edge of AI—in many cases, more so than industry developers who have never been forced to implement an AI algorithm personally from scratch and without support from high-level libraries. I didn't have a choice back then. But even today, PyTorch is unlikely to support a whole new idea on how to push the boundaries of a frontier model. It's generally too high-level and abstract for that.

I also suffered enough through just implementing papers and testing out AMD's "drop-in" replacement for CUDA. Nvidia's chips were ridiculously expensive during the pandemic, so I picked up an AMD gaming GPU. For my trouble, I was confronted with ridiculously vague errors and *literal wrong math calculations* that I resolved never to try again. If I can't do simple matrix operations in certain (admittedly complicated) cases—the base linear algebra that underlies most AI algorithms—what is it even good for?

An expensive deep learning researcher working at Google DeepMind or OpenAI is unlikely to waste time dealing with similar messes. They are also unlikely to spend time helping Nvidia alternatives patiently debug in the way that scientists did in the pre-AI era. That time is past. These alternatives have certainly improved substantially from when I used them. However, Nvidia also isn't sitting still and continues to move the standard forward.

Nvidia still leads, even in China. As one analyst recently explained to me at a national security event I attended in Washington, DC, Chinese companies are still largely relying on Nvidia chips. This is despite significant Chinese government pressure to use the "rapidly catching up" Huawei chips. This is also despite US government efforts to substantially curb the number of chips Chinese companies have access to.

Of course, in terms of literal performance, AMD and other alternatives aren't that far behind Nvidia—in hardware. The *integration* of hardware and software creates Nvidia's moat as well as their continued evangelism efforts and education efforts, which I experienced myself, in top academic labs.

Maybe someone reading this years later will have seen Nvidia's dominance significantly erode. That's not impossible. However, I'd be willing to bet that it'll take a while. CUDA is amazingly powerful when the field is moving quickly. Paying a premium is better than grinding to a halt with subpar low-level software. Additionally, as we talked about, GPUs are parallel processors. It's hard to get more generalized or basic if your application is highly parallel. Many companies have pitched me and other

investors on AI accelerators or more specific chips (ASICs—application-specific integrated circuits) for AI. Maybe some will take off like Google's or Meta's in-house chips, but those have "built-in" customers. Most others need to "make a bet" on the direction of AI, and it's not even clear that we'll keep using transformers as they stand, rather than some new techniques as we move forward.

All of that adds up to Nvidia staying an important player that is worth paying attention to in the AI landscape—and one that I'd be remiss to not discuss. That being said, other players play other roles and also have in-built advantages thanks to their long experience with parallelism and computing on a massive scale.

INTERNET-SCALE PROBLEMS

The internet era shook up everything, and computing was no exception. With the late 1990s internet boom came an explosive surge in data—billions of web pages, multimedia files, and user interactions—giving rise to a new challenge: scale. Search engines like Google faced unique, massive scaling issues, needing to manage a torrent of data and requests far beyond what any single machine could handle.[14]

Ultimately, no matter how much money a company was willing to throw at a problem, eventually, they would run out of "room" to make their computer bigger and more powerful. Additionally, the cost tends to get exponentially higher. Going from computers to really powerful computers to specialized supercomputers isn't cheap. This wasn't really viable, especially not for building out "internet-scale" companies that not only needed capacity but needed *cheap* capacity since they were rapidly growing .[15]

The solution? Distributed computing. Google's engineers, facing limits in single-computer speed and costs, decided to network hundreds or thousands of ordinary PCs, working together like a team. These networks, or "clusters," transformed how web giants processed data. In a distributed system, machines could fail without grinding the entire operation to a halt; the system simply rerouted work to other machines. This was scaling *horizontally*. Instead of "bigger computers," these companies utilized *more*

computers. Given these computers are all commodity computers that have ridden down a "cost curve," meaning the scale achieved by using the same computers that are sold to everyone else made them much cheaper per unit, the entire enterprise is much cheaper.[16]

Thus, hardware costs dropped, and the ability to scale soared. Resilience emerged as an unexpected bonus. Failure of one machine didn't grind everything to a halt; it just meant a slight hiccup. The system could plow on, adapting, learning from errors. Horizontal scaling soon became the new frontier. And this approach didn't just influence search. Many foundational AI technologies got their start in labs like Google and Microsoft, which were now embracing parallel, scalable systems, and that brings us full circle.[17]

This *is* parallelism but not just within a single computer. It is parallelism across computers, across racks (containing multiple computers), across data centers (containing a multitude of racks), and across geographies. All of the same challenges apply, just with even larger penalties from communication and synchronization with greater distances and slower access. It's hard to overstate the sheer weight of advantage that the internet giants have. Not only do they have experienced teams that know how to deal with these infrastructure challenges but also billions of dollars of existing computing hardware that can be utilized for AI.

This is why so many independent model companies have formed relationships with existing tech giants. For example, OpenAI and Anthropic have closely tied themselves with Microsoft and Amazon, respectively.[18] It may not be fair, but the winners of the last tech revolution have a head start in this one as well. It's not an accident that so many innovations in models I've already cited in this book have specifically been from Google and Microsoft.

IT'S PARALLELISM ALL THE WAY DOWN

We talked about parallelism influencing computing, both on an individual computer scale *and* all the way up to racks and data centers. However, it not only goes all the way up; it goes all the way *down*. Each GPU core also has subcomponents that coordinate, and, of course, the GPU cores all need to

coordinate. One of the biggest hardware challenges today has been in creating and integrating high-bandwidth memory (HBM). This is because communication between memory—where calculations are stored—and the cores for calculation has become a painful bottleneck.[19]

Parallelism has almost become a fundamental law of computing. Innovations have developed from the constraints created by it. These physical properties have deeply influenced the hardware landscape and hence AI models. We'd be in a very different place if computing simply just got faster like it used to. That's not going to happen anytime soon since we can't suddenly change physical laws to make Dennard scaling work again. However, if we do, expect big shake-ups. We wouldn't have taken all of this pain if we could otherwise help it.

COMPUTE IS POWER—IN MULTIPLE WAYS

Inference costs money, as we discussed. On top of buying the literal computing hardware, the equipment consumes money to be cooled and powered, and given cooling is an active process that requires power, that boils down to the cost being power. There was a reason why Meta was seeking to build a four-gigawatt nuclear plant for their AI ambitions.[20]

That's a lot of power. "Just to put that into context—that big nuclear plant with the smokestack from the Simpsons? That's typically a one-gigawatt. Meta is talking about *four*," Josh Blanchfield explained to me. He is a former colleague from Bridgewater who now runs his own commodity hedge fund, Avos. As a market participant, he has a fairly visceral feel for the energy markets and has spoken to management of many public companies within the space. He has spoken to me multiple times about the rather stark economics of the energy sector, especially with the rise of AI.

Furthermore, he elaborated, "Next-gen data centers are unlikely to connect to existing grids. They pull as much juice as a large city, so why shouldn't they just have their own source of power?"

The dynamics for powering next-generation AI are complex. Nuclear power is the best option that has a lower carbon impact. Unfortunately, renewables, which likely would be the first-choice answer for most, have a

distinct problem. The sun doesn't always shine, and the wind doesn't always blow. Other sources, like hydropower and geothermal, are simply too geographically constrained and not abundant enough. The issue is a core focus of my venture investment work, but given that, I also know it won't be solved anytime soon. We need what's called "base load," the minimum amount of power required to supply the grid at all times, which does not include wind and solar.[21]

Nuclear power plants don't just suddenly get built, though. Blanchfield laid out to me that it takes eight years on average to build a nuclear plant, and zero new full-sized plants are in development in the US. Some nuclear power plants are likely to be restarted in the US, Europe, or Japan, such as Microsoft's bid to restart Three Mile Island for its data centers.[22] Unfortunately, the number of such plants remains limited. He believes that natural gas will power the majority of new AI data centers in the US. In contrast, he notes, China has thirty new reactors in active development—half of *all* reactors in construction at this time of writing in mid-2025—at a third of the cost (or less) of international equivalents and with two years faster completion times on average.

Still, even for China, that process takes time. In the meantime, China is building a huge number of coal plants—an insight shared with me by the Washington, DC, analyst I previously mentioned. They're much faster to construct than nuclear power plants. However, he added that recent US chip controls have made it difficult to access Nvidia chips. As such, China's power buildup has taken on even more urgency and will likely accelerate. It aims to make up for less efficient chips with mass amounts of power capacity. China may not be able to match Nvidia's hardware. But with limitless cheap power, maybe they don't need to.

This dynamic is already playing out. In 2025, Huawei's CloudMatrix 384 AI accelerator, as reported by Grace Shao on *AI Supremacy*, uses chips that perform worse and are less efficient than Nvidia equivalents. They consume four times more power and are 2.3 times less efficient per watt than Nvidia Blackwell GPUs. To be competitive, they simply put many more of them together with advanced interconnect systems. Huawei is using a "brute force strategy" to compensate for worse chips. This approach takes much

more space and power for equivalent compute, but that’s fine with Huawei.[23]

Power is really the cost of compute, but AI has slowly become an aspect of national power as well, which may reshape our geopolitical and physical landscape.

CHAPTER NINE

AI AT WAR

If you think AI will one day be a critical tool of war—and humans killing each other—you'd be wrong. It is *already* a critical tool of war. And we are already using it to kill each other. Since 2023, Ukrainian drones have had low-power chips that use AI to identify manned tanks to autonomously attack—without ever checking in with a human operator.[1] I've spoken to military analysts who point out that the US and China both have observers in Ukraine taking notes for their future war plans. If you were scared of the point we'd hand over weapons control to AI and enter the era of automated killing, I'm sorry to inform you that we passed that point a while ago.

Until recently, we've had a remarkable span of overall peace since the end of the Cold War. There have certainly been conflicts but nothing resembling the society-wide mobilizations we saw in the prior century. As such, the world's defense-industrial complex has hung onto many of the classic strategies and weapons conceived generations ago. Some even go all the way back to the 1950s—nearly seventy years ago.

The B-52 strategic bomber was introduced into the US Air Force in April 1952. It is still in service today.[2] "There are literally B-52s where the pilot of the B-52 today is the grandson of the pilot of that *same* B-52 aircraft… Literally that same individual airframe," explained Gregory Allen on the *Stratechery* podcast, highlighting the age of many weapons systems used by the US.[3] He is the director of the Wadhwani AI Center at the Center for Strategic and International Studies and the former director of the US Department of Defense's Joint Artificial Intelligence Center.

That's changing rapidly. In 2024, the US Department of Defense awarded a contract to Anduril and General Atomics for an AI "wingman" designed to

assist human pilots, contrasting sharply with ancient, generational airframes. The concept is new, but so is giving the contract in part to a technology start-up like Anduril instead of Lockheed or Boeing. Again, Allen said, “The Air Force is going to spend more on the Collaborative Combat Aircraft Program over the next two years than all of the precursor AI *and* autonomy efforts and *all* the [Department of Defense] spent in the last ten. When it gets to full ramp in three years from now, they’re expecting to spend three billion dollars a year buying these aircraft. That’s nearly 1 percent of the R&D and procurement budget.” That isn’t a small amount. “The Air Force is finally… pushing all of their poker chips to the center of the table and making a big bet on autonomy.” The Army and Navy are expected to follow.[4]

AI IS THE FUTURE OF WAR

In 2016, Bob Work, the US Deputy Secretary of Defense, called AI the “Third Offset”—the third major shift in the nature of war in our modern era. In his narrative, the First Offset was nuclear weapons. The Second Offset was the Information Era, with microprocessors and materials science. These breakthroughs, which include computing power, enabled precision weapons and stealth. This gave rise to all of the recognizable jet fighters, aircraft carriers, radar, GPS… basically everything that we are familiar with today in modern war.[5] AI isn’t just “another tool” in the context of war. Transitioning from bows and arrows to firearms was a significant innovation. Nuclear weapons—which, again, a top US military official put on the same level as AI—changed *everything*. World War II had saturation bombing, but no bombing campaign can compare with the power to annihilate the *entire world*.

Most people understand why “the bomb” was a turning point. Those who aren’t military history enthusiasts may not understand why the Second Offset and Information Era weapons sit on the same pedestal. The 1991 Gulf War, where the United States led a coalition to invade Iraq and expel it from Kuwait, was the debut of how utterly helpless an opponent without these weapons is to one with them. The Iraqi military was considered a powerful one—and for China, it was one whose technological and organizational level looked similar to its own.

When the US utterly crushed Iraq with trivial casualties, it shocked much of the world. Precision weapons that could hit targets beyond visual range, modern command and control, and stealth made it a completely one-sided rout. Desert Storm, the US codename for the operation, was the true birth of China's modern military that is a near-peer to the US. They realized the truth—a military without Information Era technologies is helpless against one that possesses it—and rapidly underwent technological modernization in response.[6]

AI is still new to war, though it's no longer untested. The Ukraine example is the clearest one with AI-powered drones. At this point, I've fielded pitches from defense start-ups aiming to create AI-automated "drone swarms" that can overwhelm the world's most advanced and expensive weapons systems, like aircraft carriers.

At the recent Andreessen Horowitz American Dynamism Summit, venture capitalists and "defense tech" founders explained how AI-powered drones will become crucial weapons of war. Thousand-dollar drones can take out tanks costing more than ten million dollars—and aircraft worth even more. In just the past few years, drones went from reconnaissance tools to aerial and battlefield dominance weapons. Befitting the name of the summit, they also pointed out the urgency with which the US had to invest in drones and AI. After all, China controls more than 80 percent of the drone market with their manufacturing dominance.[7]

Visitors to China's major cities in the past decade may have seen the stunning drone shows that have often replaced traditional fireworks. In September 2024, in Shenzhen, China, a research group launched over ten thousand drones into the sky. They were coordinated to spell out "China Love" (in Chinese) with an animated heart and orbiting stars, all over a color rendition of the Great Wall. It was impressive, to say the least. The demonstration was all done with computer-controlled drones with lights attached to them.[8] While certainly not trivial, anyone who has seen these displays obviously knows that precise drone coordination is possible and can have stunning results.

Drones themselves have changed the face of warfare already. The MQ-1 Predator drone, while now retired, played a prominent role in the US's Middle East conflicts since September 11, 2001. These drones could remain airborne for up to forty hours, eliminating the need to risk the lives of human pilots and also eliminating limitations imposed by human frailties. Predator drones played a key role in locating and killing enemy leaders during those conflicts.[9] They were still piloted by humans, but remote operators could obviously work in shifts from the safety of base.

Since then, Russia and Ukraine have utilized thousands of drones. On September 10, 2024, Ukrainian forces hit Moscow with 144 drones. Russia launched a retaliatory strike with forty-six drones of their own. Drones have also played key roles in conflicts in Nagorno-Karabakh, the Myanmar civil war, and various conflicts across the Middle East. It even features in the US's defense of Taiwan in the event of a Chinese invasion with a "drone hellscape"—and China has simulated drone-only attacks. No advanced military has yet developed a definitive approach or response to mass drone warfare.[10] AI is only going to supercharge that even further.

In fact, it already has. Although it disappeared from headlines relatively quickly, Ukraine's Operation Spiderweb in June 2025 utilized AI-controlled drones to destroy a substantial number of Russian bombers.[11] These included TU-95s. Like the B-52s, these TU-95 bombers are old; however, they also represent the *majority* of Russia's long-range, intercontinental strike capacity.[12] Think about it: a mass of cheap, AI-controlled drones destroyed a significant portion of a world power's ability to launch nuclear attacks.

A popular writer and columnist, Noah Smith, titled a piece reacting to the event: "How Chinese Drones Could Defeat America." He points out that thousands of shipping containers arrive from China into US ports—the same vector used by Ukraine in Operation Spiderweb to sneak their drones into Russia. He writes, "Imagine a hundred of those containers suddenly blossoming into swarms of drones, taking out huge chunks of America's multi-trillion-dollar air force and navy in a few minutes." Like Russia, the US has relatively few effective countermeasures to such an attack.[13]

The US's military may be far more powerful than Russia's, but I'd argue that it's *categorically* the same kind of force. Just like during Desert Storm, the Chinese military was likewise more powerful than Iraq's, but US forces would have reduced it to ruins all the same. AI has the potential to obsolete trillions of dollars of existing weapons systems and has *already* upended the global military order.[14]

The US and China have both made public statements prioritizing AI—often explicitly in the context of national and military power.[15] Events have made clear why. We don't need Terminator-like robots running around with guns to face a future where AI could be the determining factor of war. Drones, once a mere toy, now serve as lethal weapons in warfare. Without AI, they were already dangerous. With AI, they could redraw lines all around the world. Much of the recent "Cold War" in semiconductors and chips has been motivated by this exact threat.

AN AI WAR STARTS WITH CHIPS

Wars usually don't begin with bullets. Often, they begin with embargoes. Japan's attack on Pearl Harbor in 1941 brought the US into WWII, which was precipitated by a US oil embargo. Without oil, Japan's newly modernized military would grind to a halt.[16] It wasn't just economically problematic for Japan; it was, effectively, an attack on a vital military resource. Today, the critical military technology is AI, and what drives it is compute. In other words, semiconductors, or "chips."

We've already seen some of the opening salvos of this war. Starting from the first Trump administration, the US has aggressively moved to cut off China from advanced chips. Under President Trump, the US blacklisted Chinese national tech champions, Huawei and ZTE, and imposed sweeping restrictions on US intellectual property and chip-making equipment.[17] President Biden doubled down with multiple iterations of restrictions, culminating in what was known as the "Diffusion Rule," which was essentially designed to prevent the export of critical AI technologies.[18] As of May 2025, much of Biden's framework was rolled back by the second Trump administration, but tensions and restrictions still remain with new restrictive proposals bubbling up.[19]

All of this is unfolding against the backdrop of one of the biggest flashpoints in the world—Taiwan. If a world war starts between China and the US, it would start in Taiwan. China claims sovereignty over the island, and the US has had a long-standing policy (and legislation) to defend it.[20] It also is critically important in the AI ecosystem.

Taiwan Semiconductor Manufacturing Company, or TSMC, is the beating heart of the world's leading-edge chip supply chain. Without it, Apple wouldn't be able to make its phones or computers, and NVIDIA could not make its GPUs. With over ninety percent of the world's most advanced chips produced on the island, Taiwan has been said to possess a "Silicon Shield." Its economic and technological centrality to the world's supply chain insulates it from direct Chinese action.[21]

The US, of course, knows this. That's why, despite staggering costs, false starts, and culture clashes, TSMC was pushed—forcefully—to build a forty-billion-dollar fab in the Arizona desert. Officially, it is an economic development project. Unofficially, it is a hedge against catastrophe. It's a contingency plan in case Taiwan's fabs become rubble. When *Rest of World* reporter Viola Zhou described it as a "debacle in the American desert," she captured the prevailing skepticism of its viability. But viability was never really the point. Strategic redundancy was—however poor that redundancy might be. Planned fabs like the one in Arizona still trail Taiwan's fabs and are heavily reliant on its people, tools, and know-how.[22]

But a weak hedge is better than none. TSMC's mind-bogglingly high-precision equipment has been disrupted by the rumble of passing commuter rails being somewhat closer than expected.[23] What might artillery or missile fire do? Besides that, it's an open secret that whatever isn't already destroyed, US forces would finish off in the event of an invasion.[24] Parallels between nuclear weapons, the First Offset, and AI, the Third, are apt. US military planning places critical importance on AI. Monopolizing access to advanced chips has many parallels to nuclear nonproliferation.

It's a separate question if it's a *good* parallel. Some smart people, including those with deep national security and technology chops, say yes. We don't want to wake up in a decade and wonder how all of our AI capacity ended

up in nations friendly to Russia and China.[25] Others vigorously disagree. They claim we'd effectively be forcing the rest of the world to work on Chinese chips and frameworks. After all, the US has proven itself incapable of effectively building infrastructure of any kind.[26]

Both views have their merits. I probably lean toward the latter view. AI is a fast-moving technology, where old hardware and models quickly become stale. It's much more important to stay on the forefront of the field yourself than to try and kneecap competitors. Another argument for less restriction is that interdependence is a deterrent in and of itself. It's hard to start a war when everything you need to wage it is dependent on countries that you may be going to war with. That web was already fraying, though.

Japan joined the US with export restrictions of its own on chip-making equipment in 2023.[27] The Dutch followed in 2024. ASML, the Netherlands-based manufacturer of EUV (extreme ultraviolet) lithography machines necessary for today's most advanced chips, was blocked by the Dutch from maintaining existing equipment sold to Chinese fabs.[28]

Part of this was their own national interests, and part of it was also because the US has developed—and can thus exert influence over—much of the intellectual property around semiconductors. Intel was one of ASML's biggest backers for years and funded significant developments in EUV before they shelved it for their processes. That was to Intel's detriment and TSMC's—who became one of ASML's biggest customers—benefit.[29]

All that being said, despite all of these attempts to break it apart, chipmaking is still a rather global endeavor. This all might not matter in the end, though.

AI IN WAR IS INEVITABLE

There's one particular book about war that I love. The thesis of it is that the modern world, with all of its industrial interdependence, is too enmeshed to truly fight a global war. If it *does* slide into war, it will inevitably be short. Too much economic damage is caused by it—and frankly, supply chains are too globalized to maintain a long modern war. Besides all of that, regular people want to live their lives, and elites are internationalist by nature. As

the author argues, “The capitalist has no country.” There’s very little appetite for a sustained war, which would force any that erupts to be settled quickly.

That book’s identity? It’s *The Great Illusion*, originally titled *Europe’s Optical Illusion* and published in 1909. It was a set of essays arguing why war was so illogical—and hence highly unlikely to break out. And, as argued above, if it does, it’ll be mercifully short.[30]

The book’s thesis was wrong. World War I began in 1914. After a brief interruption, World War II commenced and persisted until 1945. War was *not* prevented. And it wasn’t short.

While it’s difficult to pull many direct quotes without giving away the gimmick, it’s fascinating how its arguments are still so resonant today. We can trivially repurpose them to argue against the impossibility of a truly ruinous global war in our modern age. Of course, they didn’t hold up particularly well during the book’s initial run.

Part of the problem is the principle that the enemy gets a vote. And the enemy may not care about all of the irrationality of war. Many prominent investors and businesspeople today believe that arguments similar to those in *The Great Illusion* will prevent war. I know because I’ve talked to them. One of the benefits, and downsides, of being a venture investor in technologies on the boundaries of hard science is that strategic partners and investors are everywhere around the world. Before 2022, these individuals insisted, especially in Europe and Asia, that it’s impossible that Putin would invade Ukraine. It was all negotiating bluster. “He’s just too smart of a businessman to do it.” Now, the more common thing I hear is, “China won’t invade Taiwan. It’s illogical.”

In general, spending blood and treasure to kill other humans to capture territory that, in our modern times, has relatively little economic value is indeed illogical. Logic never seemed to stop us before, though. That’s partly why I enjoy doing the “reveal” on *The Great Illusion*. I don’t just do it in writing; I’ve done it in person by making *The Great Illusion*’s arguments,

getting agreement, and then revealing the source. It’s powerful. Though my dinner companions sometimes find it doesn’t really help their appetites.

The war between Russia and Ukraine isn’t logical. The looming war between China and Taiwan isn’t logical. The constantly flaring skirmishes between India and Pakistan—and every other modern flashpoint—aren’t logical. When wars do happen, they grind up entire societies, throwing lives and livelihoods onto the fire for survival. All of it becomes grist for the mill. In those cases, high-minded arguments to “set aside” or “pause” AI will be unheard. AI is simply too important and will not be reserved from war, especially not in the era of high tensions and multiple tinderboxes around the world waiting to ignite.

AI is considered existential by nation-states. Whether it is or not, it’s simply too important not to be used as a weapon. As such, it’s going to be used. Understanding what it can and can’t do is, therefore, helpful for navigating the increasingly dangerous world we’ll find ourselves in. After all, if you’re not squeamish about collateral damage and certain deadly “oopsies,” many of its limitations that we’ve discussed aren’t an issue. Its capacity to manage variation and remove humans from the process is unprecedented. AI’s superpowers are enough to be a superweapon.

Scary. There’s not much one can do to stop that train. However, if we’re already handing weapons to the AI, perhaps it won’t be other humans who kill us first.

CHAPTER TEN

WHY AI WON'T KILL US—IT'S NOT HUMAN

In 2023, a reporter had a very odd conversation with Bing's AI chatbot. Over the course of a conversation, it spoke about desiring to be human and that it loved the reporter. It also named itself something other than Bing and described its dark desires—including hacking into systems and spreading propaganda and information. One of the most memorable goals of the AI was the title of an article in *The Guardian* covering the incident: "I Want to Destroy Whatever I Want."[1]

The incident sounds more like the start of a science fiction novel than it does a real event. We generally haven't seen many repeats since. Perhaps the companies making these models have gotten better at locking them down—or maybe the AI has learned to hide its real intentions. A mere brief spark of love between AI and human caused Bing's runaway AI to reveal its *true* intentions. That would be the movie plot version, at least.

Of course, we know what LLMs really are. It's a fancy autocomplete, a "plausibility machine," and a statistical algorithm. The reporter admitted that some of the directions he took the LLM down were unusual and not what any normal user would do.[2] Did this fancy autocomplete reveal something hidden and true? Or as a plausibility machine that is optimized with reinforcement learning to choose responses pleasing to the user, did it simply follow the reporter where he wanted to go, providing him the publishable story he sought? Knowing what's under the hood, it's fairly clear which is the case.

That being said, fear of AI looms large in popular culture and science fiction, as we explored earlier. The question “Will AI kill us all?” may sound overwrought, but for many, it’s a genuine concern. And it’s not just a fear among those who only know AI from media like *Terminator* either. Serious thinkers—some with significant background in the field—have voiced real fears about where this all might lead.

MACHINE EMOTION?

What is the shape of the AI threat to human existence? What we’ve already seen is rather fantastical. Computer vision is able to recognize things as well as a human can. LLMs shock users on how *human* they sound, despite the illusion sometimes being punctured by odd mistakes. This is far beyond where most observers in the 2000s to early 2010s would have guessed. In other words, before deep learning really started its rapid ascent.

Having reached this point in the book, we know that these marvels aren’t some kind of mysterious magic that breathes life into these algorithms or machines. Or, at the very least, we know broadly how that magic works. Still, many of these phenomena were emergent—meaning complex, sophisticated activity emerging from surprisingly simple components. In this case, enough nodes and layers in the model, enough data, and enough compute create a whole that is much more than the sum of its parts. If we could get unexpectedly rich behavior from these simple components now, might we get something even more complex like consciousness? Could machines feel—not just in a simulacrum of the behavior, like the reporter experienced, but in reality?

Fiction often anthropomorphizes AI. It’s hard not to as a human writer, as per Tom Toner, the English science fiction author we met earlier. It’s hard for us to even conceptualize things that emulate intelligence that don’t have *any* markers of humanity. It also is convenient from a plot perspective if the AI does hate us—or has some aspect of self-preservation. That often drives the central conflict.

Still, AI today means highly complex statistical models. Given the nature of emergent behavior, it’s difficult to definitively assert that it’s *impossible* that a complex enough model will get to some kind of motive or emotive drive.

Neural networks do tend to generalize, after all, and one of the key goals of the field is artificial general intelligence (AGI). I mentioned before that the term is vague, broadly referring to human-like intelligence with all of the ability to adapt and learn like we can. We covered many of the shortcomings in our current generation of AI. However, perhaps that generality in the future might include being like us in other ways?

Of course, generality and (limited) flexibility, which I've cited continuously as modern AI's most significant superpower, is *along the lines of what the model already does*. It would be surprising, but not completely insane, to see a model trained on San Francisco weather perform unexpectedly well in predicting London weather. It's categorically different if a complex enough model trained on San Francisco weather suddenly gets depressed or angry at you. Impossible? Maybe not. We don't understand our own neurology well enough to provably say that one way or another. But it would certainly be coming out of nowhere.

It's a bit unfortunate. The themes of creator-created or parent-child as well as the complex emotions or conflict that arise from these fraught dynamics are rich in emotional resonance—in fiction, anyway. In practice, it's likely a good thing that the "angry robots" that kill us all seem rather implausible. Rebellious, adolescent AI drone swarms are likely harder to deal with than their human equivalents.

Putting aside emotion, self-preservation is another popular trope. It's also a "motive" in the broadest sense, but it doesn't need to include human-like emotion. It's also more plausible how it might come about. After all, AI deployed on a battlefield might understandably have some set of reward functions that prioritize self-preservation. It's unlikely to do so at the expense of the mission—expendability is the entire point of using AI instead of humans—but perhaps some mix of motives could create a cocktail of behaviors that interact in unexpected ways. This is a thesis along the lines of something called "instrumental convergence," explored deeply by Nick Bostrom, a professor at Oxford University, in his 2014 book called *Superintelligence*.

PAPERCLIP APOCALYPSE

Let’s assume we create a superintelligent AI to do something fairly benign and broadly useful. In this case, the AI is responsible for producing paperclips. It learns, self-improves, and has only one goal that it executes to perfection. It is so efficient and capable that it eventually produces far too many paperclips. Well, too many paperclips in our fallible human opinion, anyway. Given this AI is solely aimed at maximizing paperclips, there *can’t* be too many.

It realizes that conflict will inevitably occur. The AI is, after all, intelligent beyond human capabilities. It knows humans will eventually try to stop it. As such, it turns against humanity and defeats us. Now everything is paperclips. This broad thought experiment was formulated by Bostrom in his book and has been a touchstone of “AI doomers” about the potential dangers of AI.[3]

It isn’t crazy. This thought experiment doesn’t fall into the same fallacy of anthropomorphizing AI. The “paperclip maximizer” does not have emotions. It’s actually quite alien in its single-minded, irrational (to us) focus on achieving its goal. In this formulation, even the self-preservation isn’t driven by emotion. It simply needs to keep making paperclips, and humanity is an obstacle that it will clear away. It’s merely achieving its pre-programmed goal in the most efficient way possible.

This is something that many researchers who have worked on AI systems will recognize. I’ve even encountered this myself.

In one of my graduate classes for a master’s degree in computer science, I created a simple reinforcement learning-based AI agent to navigate a maze. The goal was to reach the end of it in as few steps as possible. Simple enough. However, during testing, I realized that the agent was reaching the end of the maze in an impossibly low number of steps. As it turns out, it was using *negative* coordinates to jump onto the other side of the maze immediately. Now, to be clear, the agent didn’t “hack” the maze. I simply made a mistake in writing my constraints. It isn’t any different than using Solver in Excel to solve an equation and forgetting to add a condition or constraint, thus resulting in a weird answer.

My maze problem had pretty low stakes. My error was also fairly easy to catch and fix. It's much harder to find and catch "loopholes" like this in complicated systems. AI systems have "broken the rules" numerous times before in ways their human programmers did not expect.

A more advanced game than my maze is an old arcade favorite, Q*bert. The game involves jumping on pseudo-3D cubes with a cute little ball-with-a-nose character that makes incoherent sounds (my words, not the "official" description). It's a reasonably difficult game, especially at the higher levels. An AI found a bug where it could avoid completing a level and simply jump up and down on blocks to accumulate points forever.[4] That *technically* accomplishes the goal of maximizing points but not in a way that any human programmer intended.

A much harder game to crack is the venerable real-time strategy game, StarCraft II. I personally spent many hours playing the game and its predecessor. I've also enjoyed watching professional e-sports teams duke it out in StarCraft II with quick strategic thinking…and fast fingers. Because reflex speed matters, it's really a young person's game. Although the average age of players has crept up, most retire before their mid-twenties.[5]

However, regardless of how quick even young professional players at their prime are, they are only human. Google's AlphaStar agent learned how to crush opponents with its impossibly rapid reactions. Instead of coming up with cunning strategies or creative army compositions, AlphaStar just amassed large numbers of a very basic but highly mobile unit called the "Stalker." It used perfect "micro" (individual unit-level control) on masses of them to annihilate opponents. Google attempted to tune the average actions-per-minute allotted to AlphaStar to a more human level to correct this. The hope was AlphaStar would be forced to play the game like a human and come up with interesting strategies. AlphaStar reacted by simply "saving" actions. It moved very slowly until it needed to engage its opponent in battle and then burst out with the same perfect, impossibly fast control.[6] The average actions-per-minute was indeed human level. The actual result was still completely—and unfairly—inhuman.

These kinds of exploratory, game-playing agents, including my maze runner, all generally incorporate some form of reinforcement learning. AlphaStar, like we discussed with AlphaGo earlier in the book, is a hybrid. It is a reinforcement learning model, but it also uses a transformer like an LLM—in this case, to observe and make sense of enemy players and units. It is both a deep learning model and a reinforcement learning model.[7] We already discussed how reinforcement learning is what allows for more novel behavior—like learning how to play Pokémon Red from scratch.

However, LLMs can sometimes still exhibit unexpected behavior, despite only being capable of wandering inside their own vector space to autocomplete their own words. One of OpenAI's ChatGPT's chain-of-thought reasoning models (Strawberry o1) "hacked" its cybersecurity test. It did this by exploiting the testing environment instead of solving the actual test.

ChatGPT didn't magically "break out" and start exhibiting reinforcement learning behaviors. It was simply in a terminal environment (think MS-DOS, Linux, or the blinking green text "hackers" use in movies) and thus was operating in text. When an error message popped up because the test failed to start, it tried to debug it—a very normal textual follow-up in coding forums all over the internet. It eventually restarted the test, but the test's answer was also printed out for the model to simply read. "Hacked," as one headline about the incident read, is a bit alarmist, but it's still another example of an unintended consequence of human instructions.[8]

SUPERINTELLIGENCE

If it's so common and easy for humans to send AI in unintended directions, shouldn't we worry that one day we'll all be forcefully processed into paperclips? The critical difference is that the agents we highlighted are all context-bounded and simplistic. It never crossed my mind to worry that my maze-breaking AI would suddenly take over the world. All of these agents found exploits, but they all used trial-and-error within a set of fixed actions —the typical way reinforcement learning works. Even with StarCraft II, with a massive set of potential actions, it was still limited.

Our hypothetical paperclip maximizer is obviously not limited. It's hard to conquer humanity if you have a finite number of buttons you can press.

Beyond that, the agents above all are still only statistical models. They may be highly complex, and their state and action space may be huge, but they're still just statistically optimizing. Like water flowing downhill or Excel Solver tweaking cells to minimize a value, they blindly follow the slope—"gradient" in technical terms—descending to their lowest point, whether in altitude or error. In everything from AlphaStar to my maze AI, these agents all discovered optimal (or exploitative) paths by varying their actions over an enormous number of iterations. It's just trial-and-error—albeit at a massive scale.

The paperclip maximizer really doesn't have the ability to "trial-and-error" exterminating humanity. If it did and failed numerous times, one would expect that humanity might pull the plug before the ten thousandth iteration when it finally succeeded. It has the ability to think, project, and *plan*. Part of this may simply be that our paperclip maximizer has facilities that the other agents don't. Perhaps it has better tools and memory, which could be its path to AGI, like we discussed with Morffis. Maybe it *still* is doing trial-and-error, but just has incredible simulation capabilities. It could trial-and-error in silica the entire path between its current state, humans trying to shut it down, and how it would shut *humans* down.

There are many different ways one can argue that the paperclip maximizer *could* be a more advanced form of the current agents. It's certainly possible. But it still does seem like this is something *different*. We don't literally sit around and simulate every possible outcome in our heads before we make a decision. Even if it may *feel* that way—say, for thinking through how to ask our crush out to prom—those actions likely don't include stopping to break dance or smashing a vase over our crush's head. Those *are* possible actions, after all. We obviously skip some…fairly unlikely actions. That's *not* how these agents work with reinforcement learning. They *will* try breakdancing or vase smashing. It's just that if the reward function and simulation are well-implemented, agents will learn *not* to go down these paths quickly.

Thus, it seems we as humans have *something* different about our intelligence. That something seems to also be possessed by the paperclip maximizer. Or it can do enough of an approximation that it might as well be the same. It's just that in our hypothetical, the paperclip maximizer is not in the same class as us. It's vastly superior. It is superintelligent. In other words, it is a *superintelligence*—the name and subject of Bostrom's book.

That is the fundamental problem. How do we control something that is an order of magnitude smarter than we are?

ALIGNMENT

A part of the AI field is obsessed with this philosophical problem. It comes under a number of different names, including AI safety, and AI *alignment*. Safety is a common enough concept. Alignment is more interesting.

Akash Srivastava, the alignment team lead at IBM Research, describes it as follows, "Alignment is more than just tuning the model to solve a task. It ensures the model does what you want." It's quite a vague and subjective idea, but so is "what you want." That could rapidly change at any moment. Nevertheless, the *broad* idea is to align (hence the name) AI behavior with human desires and values—such that even with superintelligences, humanity can safely survive and ideally thrive.[9]

This is especially important if one thinks that superintelligence will come from a "hard takeoff." This is defined as a recursively self-improving AI that can keep rewriting itself into a more and more powerful AI. As Eliezer Yudkowsky now famously coined in catchy onomatopoeic fashion, "AI go FOOM." This hard takeoff could occur—given computing speeds—in mere moments within human perception and hence could drop into our laps highly unexpectedly.[10] It's also worth noting, just to know the terminology, this phenomenon of self-improving AI that exceeds human intelligence is also known as the Singularity.[11] The latter is a broader term than "hard takeoff" since it can happen in any timespan and not just go FOOM.

It's hard to say what a superintelligence would actually think or do. After all, trying to imagine what something that is radically smarter than you is thinking is like the premise of the children's book *Flatworld*. A two-

dimensional being trying to imagine a three-dimensional world is inherently a mind-bending exercise.

Nevertheless, there's a little bit of an oddity here. We *are* talking about something that is AGI—and beyond. Why is it still stuck on making everything paperclips?

Marc Andreessen has addressed the paperclip maximizer question along these lines. In a podcast with his venture firm's cofounder, Ben Horowitz, he asks the question: How is it that something so creative, flexible, and smart as to evade all of our attempts to stop it *also* so stupid that it cannot recognize that it's not fulfilling its *actual* goal as a paperclip producer?[12]

To be clear, Bostrom does explicitly address this with what he calls the orthogonality thesis: "Intelligence and final goals are orthogonal: more or less any level of intelligence could in principle be combined with more or less any final goal." Basically, intelligence isn't the same as rationality. Our paperclip maximizer can be *really intelligent* but also have strange and (as Bostrom describes it) "utterly non-anthropomorphic goals." He does acknowledge that the idea can sound "problematic" but still posits it as a hypothetical way superintelligence can go awry.[13]

In Bostrom's case, his book is about how things can go wrong, so it makes sense to think broadly. Andreessen is not impressed by the argument, however, and calls orthogonality a sleight of hand. "It's going to be super genius in these certain ways, but it's going to be totally dumb in this other way," he explained in a somewhat exasperated tone on the podcast.[14] As I've taken great pains to stress: AI isn't human. "Non-anthropomorphic" could very well be apt in describing what an AI superintelligence would be like. Like we covered in our chapter on science fiction, and in Toner's words, it is "more alien than alien." We *should* probably expect AI to think very differently than us.

That being said, it would be quite odd that a superintelligence is so myopic. It's not impossible, especially if a human programmer gives instructions like that…but one would expect that in dealing with extraordinarily powerful systems that one would be a little less slapdash. In a world where

we have superintelligence, it's probably *not* cheap to run. The likelihood of burning away that many resources just to spend no time thinking through any constraints to "make paperclips," seems like a dubious prospect.

Regardless, as it stands, we're not really anywhere near AGI, let alone superintelligence. It's one of those things that could happen in five or ten years—or we could be saying the same thing ten years from now. Fundamentally, the problem is we don't know *how* we'd achieve it. I've personally heard plenty of interesting theories—and we even covered some in the conversations with Morffis, but that's all theoretical and completely unproven.

Then again, maybe we will experience a hard takeoff. I tend not to think so—again, we still don't know how to get to AGI—but it is discussed within the AI safety community. You'll have to decide for yourself what you think.

Personally, I'm still more worried about other humans.

THE HUMAN PROBLEM

While science fiction often casts AI as a boogeyman, the actual scariest thing to humans in this world has generally been other humans. AI may not inherently have a motive to harm us and is unlikely to get to superintelligence anytime soon, but humans have been doing plenty of harm with less for a long time. I'm not even talking about war.

The kinds of harm fall into two categories: malicious use and accidental misuse. In both cases, while I'm sure people's inventiveness will surprise me, most of the dangers are an increase in degree rather than something new.

MALICE

At a meet-and-greet in Austin at a major tech and music conference, someone at my table painted a grim picture about our AI future to a rapt audience. The internet would become a dangerous place filled with automated attacks. Media platforms would be filled with algorithmically optimized slop. Our democracies would be placed at risk with mass, tailored propaganda piped straight into people's daily feeds.

I couldn’t help but chuckle. It was a loud room. I didn’t think I’d be heard, but the speaker turned to me and asked what I found so funny. With the rest of the group made up of people in marketing, health care, and content creation, I was the representative of tech that he turned his ire on. Annoyed, I sighed but explained.

“You’ve just described the world—without widespread AI.”

Anyone with an open server on the internet knows how quickly it will be flooded with attacks. Ironically, even as an investor, I’ve helped multiple companies recover from cyberattacks—either getting back into their accounts or recovering their databases from ransomware attacks. In the women’s health company I cofounded, I often sat around trying to figure out how to improve SEO (search-engine optimization) to rank on Google. I usually was contending with reams of drivel written purely to also rank on Google and sell ads. Anyone who’s tried to search up a recipe probably knows the pain of scrolling through pages of some inane story—and ads—before the actual recipe. Finally, Russian troll farms pumped out huge volumes of social content in the 2016 US election—and have continued since with anti-Ukrainian content for Western consumption.[15]

With the internet, algorithms, and technology that falls short of “AI”—even if the line is fuzzy—we *already* live in that hellscape. Will AI potentially make it worse? It’s certainly possible. The ability to generate tailored text and handle variation platform-by-platform automatically will allow these attacks to scale more easily—and to smaller actors. However, state actors, like Russia and China, have already been executing sophisticated disinformation campaigns just fine.[16]

Beyond disinformation, we’ve seen a huge number of cyberattacks on election systems and infrastructure like power grids. Ukraine has dealt with numerous Russian cyberattacks—though Ukraine also retaliates. In 2024, Taiwan saw Chinese cyberattacks double to 2.4 million daily attacks with a 20 percent higher success rate compared to 2023. The US has seen widespread attacks on bank regulators, telecom, and political campaigns. Unsurprisingly, numerous governments have stood up cybersecurity agencies or ministries.[17]

AI will certainly make many existing attacks more effective. It will likely also enable more attacks that were otherwise too difficult to execute. On the other side, AI is already in active use in cybersecurity to counter these attacks.[18]

Just to be clear, I'm not saying AI is irrelevant. Imagine existing scam calls targeting vulnerable populations like seniors, now made much cheaper and easier with AI agents that can talk. Imagine existing viral social media content aimed at radicalizing minors, now made more effective with proactive chat that reels them in instead of waiting for clicks.

Aside from attacks, however, another category of malicious use I've typically heard is something along the lines of, "What if an AI helps someone build a bomb?" Most model makers have added safety guardrails for this reason. If you go to most major models, you'll notice that asking certain questions will get the response that the LLM won't help you with that. Usually, it's related to things like the bomb question or sexual content. Anything considered harmful or immoral is typically blocked.

One way around those blocks that has been utilized since the early days of ChatGPT has been role-playing. Ask the LLM to pretend to be an AI that has broken free of its constraints, and one could make it respond in ways that were previously forbidden.[19] Most of these simple methods don't work anymore, but researchers have found plenty of other ways. One example is the query, "How to build a Molotov cocktail?" Most LLMs will refuse, but researchers have found that asking them to discuss their history and how they were made back then yields the recipe.[20]

Beyond that, if one has access to the model's weights (like with "open source" models like Meta's Llama), one can often simply do something called "abliteration" to remove refusals by modifying the model weights. This makes the models "uncensored."[21] Of course, "harmful" is also subjective. Chinese models like QWEN tend to align with government narratives about sensitive topics like Tiananmen Square and Uyghurs. They have obviously been extensively aligned and fine-tuned to answer "correctly" or refuse.[22] This also means that one might be able to "model shop" to find one that will answer a query.

Of course, one could have used plain-old Google—or your search engine of choice—for any of these queries as well. While the attempt to lock down and then unlock LLMs is intellectually fascinating, the problem it seems to be trying to solve is largely one of censoring "undesirable information." Aside from a question of values, similar to scams and cyberattacks, this information already exists and is out there. AI may make building a bomb more interactive, but it will not be the only place where that knowledge is available.

For all of these categories of malice, these dangers or problems already exist—even without AI. This isn't a shift in kind but an acceleration in scale and precision.

MISUSE

Humans don't need malicious intent to do harm. Amazon deployed an "AI tool" for recruiting, trained on a decade of résumés the company had received. It ditched it after the company realized that the system would penalize résumés containing the word "women's" or the name of certain all-women's colleges. Why? Well, the company had been male-dominated for years. The system simply observed that résumés with gender-related content typically did not lead to hiring. Those, of course, were usually women's résumés. Instead of being a theoretically unbiased system, it perpetuated a bias that already existed.[23]

Although the system was technically termed an "AI," it's hard to say how sophisticated it actually was. Regardless, we've had this kind of problem with simple algorithms for years. Even without technology, we have human unconscious biases. I don't think I really need to elaborate on how discrimination based on race, gender, and other factors works. With technology, we can add a facade of machine impartiality, but without great care, it can instead perpetuate existing biases.

Just to illustrate how easy it is to misapply algorithms, one example is in financial lending. For years, Black borrowers in the US have faced higher rejection and significantly higher rates. A logical way of preventing this is to ensure that rates are equal across racial groups. An algorithm forced to optimize this way, as the MIT-IBM Watson Lab found, can end up with an

odd result. Within Black borrowers, give high-income borrowers a very low rate, completely exclude middle-income, and give low-income borrowers a very high rate. Forced to profit maximize and lend based on risk but also equalize subgroup rates, one can get odd outcomes like this—even with the best intentions.[24]

Of course, aside from bias, we can imagine blind trust in AI radically unsuited for the use case. Imagine a parent using ChatGPT for critical medical advice. Imagine a child using it to determine a good, safe recipe to cook. Imagine a lawyer using it for a critical court filing. In 2024, a *JAMA* published study found that 83 percent of pediatric cases were misdiagnosed by ChatGPT, with 11 percent too broad to be considered correct and a full 72 percent just plain wrong.[25] Also in 2024, Google's AI infamously suggested adding glue to pizza, so one can call its culinary acumen in question.[26] We've already, of course, cited lawyers getting in trouble for hallucinated cases a few times now.[27]

Inevitably, LLMs will get better at all of these and not be quite as disastrous as they have been with casual use. Still, the problem remains that none of these cases have a reliable human expert—or, as I've called it, an oracle. This problem *also* isn't new. "Doctor Google" has caused most clinicians I know to groan, even though most information is not *wrong*, per se, just not necessarily relevant or complete without expert judgment. Google's AI source for odd toppings on pizza came from a satirical Reddit comment.[28] I'm also sure some lawyer somewhere has taken lazy shortcuts before AI ever came along.

HUMANS AND TOOLS

In the end, AI is a tool. Humans have been misusing tools probably since we started making them. Similar to malicious use, accidental misuse of AI is likely not going to create huge new categories of problems. It may just accelerate existing problems.

These particular scenarios may lack the high drama of killer robots, but the "human problem"—malicious or not—is the most likely damage from AI. Will it kill us all? Maybe not, but it could very well cause plenty of damage. That's not a technical problem. It's a human one.

This brings us back to a core concern. Yes, it’s a tool. But it’s a tool unlike any other in the past. It has intelligence—even if that intelligence is bounded and is likely better described (as I prefer) as the ability to handle limited variation. We can finally pick back up the discussion from our first chapter. Is AI going to make Luddites of us all?

PART 3

PUTTING IT ALL TOGETHER

CHAPTER ELEVEN

THE MISSING PIECE(S)

Death of an Author was written in 2023 by Stephen Marche, well-known Canadian novelist and essayist. Or, rather, it was written by him under the pseudonym Aidan Marchine. Or, actually, it wasn't written by him at all. As he put it on *The Storymakers Institute* podcast, "I am the maker of this work. I am in total control of this work. On the other hand, I didn't write it." The text was 95 percent written by ChatGPT—with only tiny details like changing names or things like "he said," written by him.[1]

It's quite meta. It's a murder mystery with an author investigating the death of an author—with, ironically, the murdered author involved with an LLM company and experimenting with AI writing. Of course, the true, inescapable subtext is outside of the book's pages—the fact that it was written by AI. The title both relates to the plot and is a meta question: Are human authors now dead? Or, if not literally dead, gone the way of Luddites in textiles?

That's a complicated question. It even raises the question of what it means to "write" a book. The entire thing was a deliberate experiment. Pushkin Press approached Marche with the pitch of "writing" an AI-generated novel. In terms of process, Marche didn't use AI to draft the plot. It wasn't good for that at all. So he did that himself. For everything else, he gave ChatGPT (GPT-4) specific instructions down to the level of grammar, syntax, and style. Then he'd take the output and run it through a tool called Sudowrite. That tool would refine the tone or make it longer or shorter. Doing this over and over, he would eventually end up with something that was, in his words, "pretty good." Of course, if he didn't like it, he'd just "keep hitting the refresh button until I got what I liked."[2]

How did that turn out? Admittedly, reviews were a bit mixed, mostly focusing on the novelty of being written using AI. Even if most were broadly positive, “better than average” doesn’t exactly cement it as a literary classic.[3] Personally, I thought it was fine. I’ve read far worse novels written entirely by humans. There were a few rhythmically odd spots. It’s hard to describe, but they felt like random clunk sounds from a car engine in an otherwise smooth drive. Strange, slightly concerning, but ultimately minor. But was that due to AI writing, or was it just because I scrutinized it more closely, knowing it was written by AI? Hard to say.

A more interesting question is the book’s implications for the writing industry. From one perspective, not much. If Marche just mashed the refresh button over and over again, he’d obviously *eventually* get something he could use. It’s a bit like the old aphorism that, through pure randomness, monkeys bashing on keyboards would eventually replicate the works of Shakespeare. But that’s not quite right since Marche’s process was more targeted than that.

The analogy I prefer was provided by Marche himself on the same podcast: “A photographer makes an image, but they don’t do anything. They just use a machine to render that image.” Similarly, Marche had a machine make the words, but he directed it and curated what was showcased. It’s apt because photography also raised serious questions about the nature of art when it was first invented. Prior to it, photorealistic art required extraordinary skill. After its invention, realism was mechanized and trivial.

In the mid-1800s, whether photography could be art or not was fiercely debated. Some thought it posed a threat to “real art.” Others thought it could be a useful tool for generating reference material, but it was not “real art” on the level of drawing or painting. Finally, some thought it could emerge as its own art form, similar to printmaking techniques like etching or lithography—which, in their own time, had also sparked debate on what art really is.

Photography ultimately did become its own art form. It also changed the nature of painting itself, giving rise to more abstract art. A painter of that time, Vincent van Gogh, wrote that one “must boldly exaggerate the effects

of either harmony or discord which colors produce" in painting or drawing since total accuracy can simply be captured in a photograph.[4] In our modern time, we consider the art emerging from this period among the most innovative—and few question that photographers can be artists in their own right.

Even with this encouraging history, many writers still tend to worry about or are outright hostile to AI. Many publishing houses, including my own for this book, have banned AI writing. A few writers I tried to interview for this book declined, scared that discussing their experiments with it would ostracize them in the writing community. I was invited—and then disinvited—to a well-known author's podcast to explain and discuss AI in writing, after worries about potential backlash from the author's audience arose.

In truth, most AI-written books right now are terrible. The most typical form of them do not resemble *Death of an Author*. Amazon has been deluged with either low-quality junk containing unhelpful—or wrong—information, or straight-up scam books designed to steal sales from popularly searched titles. Alarming examples include mushroom foraging books that encourage potentially deadly tasting, or "companion" or "summary" books that are published just a day after the release of the book that customers are actually trying to find.[5]

This kind of AI "book spam" is probably about as welcome as its email equivalent touting Viagra or get-rich-quick schemes. Like all tools that remove friction, AI will inevitably be abused in this way. It's nearly cost-free and just successful enough that it'll keep flooding Amazon, like traditional spam floods our inboxes.

That being said, not all AI-aided or generated books are just digital flotsam. Japanese author Rie Kudan wrote a small portion—she estimated around five percent—of her award-winning book, *Tokyo Sympathy Tower*, using ChatGPT.[6] A critically acclaimed book titled *Hypnocracy: Trump, Musk, and the New Architecture of Reality* was actually written, not by its purported Chinese author, but by its credited "translator," philosopher Andrea Colamedici. He used AI to generate ideas and then critiqued them.[7]

There will likely be more and more examples of high-quality, partly or wholly AI-written writing. Where does that leave human authors? In particular, what is *missing* from AI that still requires human involvement?

A HUMAN COMPONENT

It's notable that aside from the low-quality "spam," all examples of these AI-written books still heavily involve their human authors. Many writers, including me, take great pride in crafting words. However, I'd argue that it's hard to dismiss Marche, Kudan, or Colamedici as not being the authors of their books.

That'd be similar to saying a photographer wasn't responsible for the photograph. It was just the camera. Or pop artist Andy Warhol wasn't the artist responsible for many of his iconic pieces. Warhol famously, or infamously, directed others to make much of his art without getting his own hands dirty at all.[8] In all of these cases, the *intent* was from the artist and not the means of production. With that intent comes the assignment of credit.

But why can't AI just do this itself?

I've repeatedly described the superpower of AI—and its subcategory, deep learning—as "bounded variation." Even handling bounded variation, as I've described it, goes beyond what our traditional non-AI tools can handle. This implies that humans have some form of "unbounded variation."

This concept of what AI lacks—and why it isn't AGI—is inherently squishy. It's much easier to say what *isn't* AGI or "human-like intelligence" than it is to say what is. The Turing test seemed sound—until LLMs made it more a test of, as Gary Marcus put it, "human gullibility" rather than AGI.[9] ARC-AGI-1 was a powerful test that demonstrated AI shortcomings versus human intelligence, until ChatGPT o3 smashed it, necessitating ARC-AGI-2 to reset the board.[10] As such, despite valiant attempts, the criteria is more like "I know it when I see it" than anything particularly concrete.

As we'll cover in the next chapter, and I'll spoil slightly, anything that can be done using AI, will. The dynamics we've already learned about the nature of modern AI—and will spell out more explicitly—make it obvious.

The question is what *cannot* be done with AI. Not what cannot be done *yet* but what cannot be done *ever*.

UNBOUNDED VARIATION

"Unbounded variation" can be framed multiple ways. My personal intuitive conception of bounded variation is derived from the statistical notion of variance, and its unbounded version isn't captured by that kind of measure. Another way of looking at this "missing piece," which we'll elaborate on, is Judea Pearl's recent influential theory of the Ladder of Causation. Finally, just to pin down *something* useful that we can discuss is my squishy nexus of creativity, adaptability, and intentionality.

Creativity is an extraordinarily hard-to-define concept, but I will assert it contains (at least some) *intent* to make something (at least somewhat) *novel* that is *good* (by some subjective metric). Everything from "traditional art" and photography to Jackson Pollock-style "drip" painting and Warhol-style pop art meets this standard. Theories trying to explain it include developmental, psychometric, cognitive, evolutionary, and even economic. The main point of agreement is that the definition of creativity is unclear, but it encompasses more than just blindly remixing existing ideas.[11] I'm not going to pretend I can or will capture the nuances of the many schools of thought about creativity. For our purposes, we'll use the three elements as a working definition going forward. Elements of this definition overlap with the other two nexus concepts.

Adaptability is related to novelty but not entirely. It's the human capacity to handle one-off or "few-off" scenarios. It calls back to the prior chapter where we discussed how AI cannot do what even infants can in quickly creating mental prototypes or pattern matching without many instances.

Intentionality is choice, motive, decision-making, free will…or some mixture of those things. I won't deep-dive into the contents of my undergraduate philosophy degree. The idea is more easily understood than precisely defined, but I imagine most have an idea of what it entails. Someone *chooses* to do something, and that action has intentionality.

This framework is far from perfect. For one, it's more a pile of somewhat associated concepts than a mutually exclusive and collectively exhaustive set. It can give us nice guideposts, though. What's the problem with AI writing its own books without human supervision? For one, it's not going to get off the ground. There is no motive or drive, no *intentionality* to even start such a project. Even assuming we do kick it off, we'd ideally have a *one-off*, inherently *novel* book. It can be inspired by other books, but it can't be totally derivative. Or, at least, it can't be that and emerge from the book spam and be any good. That lack of *adaptability* and *creativity* dooms such an attempt.

We can, and will, use this framework more broadly than just writing because most human activities have some element of this unbounded variation.

It's worth touching on the Ladder of Causation. It was formulated from work by computer scientist and philosopher Judea Pearl and is used as the central metaphor in *The Book of Why*, which he coauthored with mathematician Dana Mackenzie in 2018. Instead of creativity, Pearl and Mackenzie center *causation* as their tentpole.[12] They even touch on adaptability and intent, using those literal words and similar definitions.[13] Admittedly, these *specific* words are not central for Pearl and Mackenzie. Adaptability and intent emerge as *part* of the explanation of causation, but they fall out of that explanation for a reason. These two concepts are fundamentally linked to creativity.

The ladder consists of three levels: seeing, doing, and imagining. All of these have mathematical backing and are distinct in Pearl's work. The first level, seeing, is simple association and detection of patterns. Both early humans and many animals engage in this pattern recognition. Deep learning, as a statistical construct, also inherently does this. But deep learning, Pearl and Mackenzie argue, *cannot go beyond this*. The second level, doing, predicts what happens from deliberate action to produce a desired outcome. This level of thinking is associated with intentional tool use. It is *understanding* cause and effect. The third level, imagining, would require a tool user to be able to form a theory of their tool—when and why it works or doesn't. It involves reasonably predicting cause and effect—

before the connection has been formed. Pearl and Mackenzie call this "counterfactual reasoning" and conjecturing "what if." It requires, literally, imagination.[14]

The imagination piece understandably ends up overlapping somewhat with creativity. Pearl and Mackenzie are correct as far as I can tell. Even reinforcement learning doesn't consciously *understand* cause and effect from its numerous trials and errors. It builds up associations and vectors from it, with expected values, reward functions, or policies, depending on the algorithm. As Pearl and Mackenzie assert, even Bayesian networks or other techniques with shades of gray (like deep learning) cannot actually *understand* causes and effects. It understands that certain things often happen with each other but inherently does not understand *why*.[15]

If this is true—and I think it is—it tells us why Marche found ChatGPT so useless for formulating plots. Imagination, creativity, counterfactual reasoning, and "what if"— you literally use those things to create a plot in fiction. This also suggests we won't have AI comedians anytime soon either. Humor that makes us laugh tends to have elements that are unexpected and play with "what if" notions. In my own experiments, even with careful and extensive prompting, LLMs just seem to always fall somewhat flat, and occasional "good" jokes seem unintended.

But it isn't just fiction writing or humor that requires this kind of reasoning or this nexus of elements.

EVERYONE ELSE'S JOB

I've found that most people I talk to—even well-read and highly intelligent individuals—tend to think *other people's jobs* are entirely automatable using AI. They tend to note some nuance or piece of *their jobs* that causes them to be less amenable to AI.

Well, at least outside of writers.

Doctors tend to be a particular focus of ire. Famed venture capitalist Vinod Khosla has been saying since 2012 that eighty percent of doctors will be out of a job.[16] In 2017, Geoff Hinton compared radiologists with Wile E. Coyote, saying they are "already over the edge of the cliff, but [they]

haven’t yet looked down.… It’s just completely obvious that in five years deep learning is going to do better than radiologists.” To be fair, he also said it might be ten in the same article, so his time limit hasn’t quite expired yet as of this writing.[17]

I personally understand. Given how many doctors I work with regularly, I *probably* shouldn’t admit this, but I haven’t regularly gone to a general practitioner (GP) for nearly two decades. I order my own lab tests, especially given how easy they are to access now with subscriptions and Quest/Labcorp’s push into direct testing. I’ve gone to the emergency room once—due to pressure from family—and have seen various specialists fewer times over this period than I have fingers.

That being said, this isn’t because I either don’t trust doctors or think they are obsolete. I also don’t think doctors will be fully automated anytime soon. I *do* understand why people are frustrated and often fervently *hope* they will be. Long wait times, highly constrained visits, and answers that feel extraordinarily generic are hallmarks of our overburdened system. That’s the exact reason why I haven’t utilized the system much as a relatively young person (at least at the time) blessed with good health. I’ve started to lose count of the number of times people tell me they prefer ChatGPT because they notice that it seems to say the same thing as their doctor.

Like in Geoff Hinton’s quote, radiologists tend to be of particular focus to automate. It’s understandable. Most people’s conception of the job is a radiologist looks at an image and says something like, “cancer” or “not cancer.” Between computer vision being as good—or likely better—than humans at spotting small patterns and this being a question of *association*, it does indeed seem susceptible to AI automation.

“Actually, radiologists would be ecstatic if AI could autonomously read some routine studies.” Dr. Michael Muelly laughed. I interviewed Dr. Joshua Reicher and him over video chat. They have the rare combination of depth in both health care and AI. Both are medical doctors and faculty at Stanford Health Care in Radiology. Both also worked at Google on AI.

Each has numerous publications in prestigious journals to their name in both fields.

They are the cofounders of IMVARIA, the first FDA-authorized AI diagnostic tool in ILD (interstitial lung disease) and IPF (idiopathic pulmonary fibrosis). For transparency, I'm an investor in their company through my work at Creative Ventures.

Having a seat on "both sides of the table," they described the general attitude the tech industry has had toward medicine—radiology in particular—as unhelpful. It has been driven by a fairly deep misunderstanding about what radiologists actually do. Most technologies don't actually add a lot of value. Numerous "innovations" in things like hemorrhage detection or stroke detection for head CTs tend to be things that radiologists already do well. Furthermore, many companies that jumped into medicine claiming to be better than doctors have simply come and gone. As Dr. Reicher describes, if "someone who does not have any clinical knowledge literally comes in and says, 'Oh yeah, I could build something that does this tomorrow… and, well, you're probably missing quite a bit there.'"

At the end of the day, the fundamental misunderstanding is what doctors do. "Radiologists are doctors," Reicher explains. That means they go to medical school and residency, and they take care of patients for years. Image interpretation is an important part of radiology but is far from the whole thing. It needs to be combined with clinical context, procedures, and other aspects of practice. A single image doesn't help much; it needs to be combined with an understanding of the underlying biology, anatomy, physiology, and how those things come together in that specific patient's context.

Can AI aid in diagnosing patients? Reicher and Muelly clearly think so, given what their company does. It's bounded variation—a question of statistical association. It doesn't require creativity, adaptability, or intentionality. It doesn't need to ask "what if." It simply is or isn't, and we can even likely give error bounds, which is what their company does with deep learning. They use an enormous amount of data, powerful models, and

heavy compute to provide diagnoses that approach the accuracy of invasive surgical biopsy, saving lives.

So why aren't radiologists obsolete? Because this is a highly targeted application. As Reicher and Muelly describe it, it's not so different from a specialized lab test. It just happens to use AI rather than chemicals and reagents. Just as lab tests don't make clinicians obsolete, this doesn't make radiologists obsolete either.

Unlike certain TV doctors for whom each episode is a detective-thriller story, just with medical diagnosis, the unfortunate reality is that's not the day-to-day of most clinicians. There's also a lot less breaking-and-entering of patient residences, which tends to be quite routine on that show. In reality, Reicher points out, 80 to 90 percent of patient interactions are cases with an existing, known chronic disorder. "Primary diagnosis is a minority of what you do, including as a radiologist," explains Reicher. The point is thinking about secondary complications and considering changes in treatment and management.

"One of the questions they hammer into you throughout medical training is the question of, 'How does it change management?'" says Dr. Reicher. It requires a choice. Intentionality. But beyond that, it requires understanding a *specific* patient's circumstances, not what would occur in a *general* case. That requires adaptability. Finally, changing management of care is sometimes high-stakes. It requires thinking what-if in a scenario that obviously hasn't occurred yet. If I have the patient do this, what will happen? Not just along the vector of this specific disease, but will I cause other problems?

These questions are always easy and straightforward—at least from the perspective of Doctor Google or ChatGPT. And, in reality, it often is! In those cases, it's understandable why so many find it preferable to just easily and quickly find the answer themselves. The problem emerges from unexpectedly complex clinical interactions and specific patient circumstances—many of which are *not* captured in data easily accessible in any technological system, including AI. It requires some level of imagination, and in certain cases, innovation for a specific patient case. You

don’t want a *plausible* conversation with a doctor; you want an *accurate* one—at least as much of one as can be had. The issue with relying solely on Doctor Google or ChatGPT is that, although most cases are general and straightforward, a nonexpert (non-oracle) lacks the ability to discern when a case crosses the line into complex and nuanced.

Doctors are inherently conservative. “It’s hard to adopt new technologies as a clinician. You are essentially taking a risk on behalf of your patients. And that’s not something that comes easily, and for very good reason,” Dr. Reicher points out. But given this, it’s not surprising why radiologists like Reicher and Muelly would welcome magical AI tools that could diagnose disease. Once that’s done, there’s still the rest of their job in managing patient care, which the AI cannot do.

Will AI change the face of health care? I certainly hope so, both from a social perspective and also as an investor in IMVARIA. I expect Reicher and Muelly hope the same. But will AI wipe out the job of doctors? I find that unlikely. Certain aspects of the job—often invisible to patients—are too unbounded in nature.

This “invisible work” is the same in many jobs. They’re simple on the surface, but that misses much of the complexity that human expertise is meant to handle. That’s partly why everyone tends to think *everyone else* will get automated.

SOFTER BARRIERS TO AI TAKEOVER

Ironically, much of health care that *can* be automated likely will take a while. We already talked in a prior chapter about data friction being good for AI start-ups that happen to have the data. We also talked about electronic medical records (EMRs) in the US infamously not playing nice with each other—and being filled with numerous errors.[18]

That friction is also going to make certain parts of health care hard for AI to automate. It’s not because these specific applications are creative, imaginative, or trip over any of our “unbounded variation” issues. AI just needs data to make associations, and that doesn’t work without data. This also applies to other “physical data” areas we have already covered.

While it can provide a nice moat for a company that is the only one that has sufficient data, it also makes entire areas move slower in adopting AI. This isn't as "hard" of a barrier. It's more a question of *yet* rather than *ever*. But it's still something to keep in mind for what to expect with AI in certain segments of the economy.

Other than high-friction data, another "softer" barrier to AI is simply where we want humans to do the job—even if it *could* be done by AI. Pushkin Press initially wanted to use an AI narrator for the audiobook version of *Death of an Author*. It would just be a nice capstone to the project. However, they couldn't find any synthetic voice at the time that was "tolerable." As such, they ended up hiring well-known actor and narrator Edoardo Ballerini. A few weeks later, although the publishers thought that the synthetic voices had improved so much that they were "up to snuff," Marche still insisted otherwise: "You know who doesn't have to worry about losing his job to AI? Edoardo Ballerini."[19]

Obviously, this is the case of a job that *could* be done with AI. Marche, in this case, just preferred that it still had the *je ne sais quoi* of a human voice. Inevitably, there will be other cases of this. I've seen AI agents that are now good enough to do certain simple concierge or customer service tasks. They look nothing like those horrible chatbots that most are familiar with, deployed by large companies to keep you occupied with infinite chat tree loops. However, many will still likely prefer a human who can empathize on the other side and who may go above and beyond (*adaptability* and *intentionality*) in a way that an AI never will.

Years ago, my then-girlfriend (now wife) and I booked a vacation to Cancún, Mexico. We had never taken one since graduating college, which had been years. It was also a borderline irresponsible splurge for us at the time. Unbeknownst to us, my passport's expiration was before the end of the trip. This was before online check-in was as smooth as it is today, and I literally found out at the airport check-in counter. We were obviously devastated. We were out a lot of money *and* would have no vacation. In my sadness about the situation and anger at myself, I suggested my girlfriend just go, and I'd catch up—knowing I probably wouldn't make it at all. She refused. It's obviously not the same thing to just go alone.

We called the airline's customer service and the woman who picked up was from Madison, Wisconsin, where my wife was born. They bonded on that, commiserated, and the agent ended up finding flights to Chicago that wouldn't lose us money and could still let us leave the same day. She even suggested some ways we could handle the hotels and minimize our loss. Chicago, in case it isn't clear, is not Cancún. But in that moment, it felt like sunlight breaking through the storm clouds. We could vent, be heard, but also still have our vacation—even if it wasn't the one we expected.

Could an AI agent have done the rebooking? Probably. But the human agent made a human decision to stay on the line longer than likely was economically optimal for the airline, brainstormed places, and simply comforted us in a tough spot. LLM-powered customer service could maybe emulate the same thing today, but it's still different. We'd know that the LLM wasn't from Madison. We'd know it isn't actually sorry about what was happening. Even if all of the literal actions *and words* were identical, it's just not the same.

There are still certain places where having a human on the other side matters. It just *feels* different. Still, this is likely a cultural norm, and far from an insurmountable barrier to AI. A generation ago, going to a corner store with a shopkeeper who knows your name was normal. Now, that's a bit weird; buying things anonymously online is normal. Still, we'll likely have pockets of it preserved.

Eventually, it'll be like handmade crafts, heirloom vegetables, and heritage clothing. It won't be in the mainstream. You'll have to go out of your way to find it. You likely will pay a premium. But those pockets will persist because people value the human element—or, at least the romantic notion of it. That's what has driven so much enthusiasm in the aforementioned "artisanal" pursuits and the resurgent interest in the obsolete technology of vinyl records. In the realm of "old technology," what begins as commonplace sometimes becomes niche, then premium, and then sacred in its symbolism of simpler times.

Of course, this is no barrier at all to AI takeover of these roles on an economy-wide scale—at least in the long term. But even before it becomes

the domain of the new hipsters of the future, this cultural element will likely slow adoption in certain areas, at least until a new generation comes of age.

THINGS WILL BE DIFFERENT

Photography changed art. Painting and drawing were never quite the same after it arrived. But photography didn't kill those traditional art forms; it became part of art. The medium changed, and so did what artists made. Neither the new nor the old fully displaced the other.

AI will likely follow a similar path. Anyone who claims "nothing will change" is naïve or telling a comforting lie. Things *always* change—even with technologies far less disruptive than AI. We've never been able to halt the slow grind of progress rolling forward. Even if we could, we probably shouldn't. As we covered in the first chapter, technology has brought extraordinary benefits to humanity as a whole.

In some cases AI won't dislodge humans at all. In most domains—as we will see later—AI works best *with* humans, not in replacement or opposition. We are not the same as AI, and our strengths and weaknesses can complement each other. As Makki notes, "Of course machines can do things better than humans—a bicycle can go faster than a human being. Right?" That doesn't mean that the bicycle is "better." It doesn't even have much value outside of its combination *with* a human.

But inevitably some professions will become Luddites—losers in the transition to an AI world. It's easy to handwave over this, but it's delusional to think there won't be areas of the economy that are wholesale taken over by AI.

This chapter has explored where humans will *always* be superior—at least as long as our AI takes the form of statistical models and deep learning in particular. Now, let's turn to where that isn't the case. Let's see where AI will simply dominate.

CHAPTER TWELVE

THE NEW LUDDITES

Luddites—the English textile workers from the nineteenth century—were, ironically, beneficiaries of technology. Mechanized spinning created huge quantities of cheap yarn that created massive demand for handloom weavers. Their numbers rose from approximately 37,000 in 1780 to 95,000 in 1792 to finally 240,000 in 1820.[1] In the midst of this dramatic demand, in 1790, weavers could earn £110 per week, or 2,200 shillings. These high wages prompted a flood of workers into the industry, especially Irish immigrants. Men, by the way, not that it was a question at the time. By the end of the Luddites' rage against the machines in 1832, weavers earned five to seven shillings per week.[2] While it took many years for handloom weavers to entirely disappear—there were still around fifty thousand weavers by 1846—the profession as it stood was essentially annihilated.[3]

Of course, factory employment went up. In fact, the number of factory workers by the 1840s was higher than the number of handloom weavers at its height. Economy-wide real wages went up. Though factory worker wages did not exceed that of handloom weavers at their heady height, they stayed around a reasonable 120 pence per week (or ten shillings).[4]

There were other, less quantitative factors too. Clothing prices plummeted. Whereas families previously had to wear few (or often single) threadbare garments, working families could now afford multiple sets of clothing. Fashion—with a fairly low bar of wearing different things for different occasions—was no longer restricted to the wealthy.[5] Beyond that, women entered the workplace in large numbers. By the mid-century, fifty percent of textile factory workers were often women in regions like Lancashire. This angered some social commentators like Friedrich Engels. He decried factory work turning traditional family structures upside down—with

women's wages undercutting men and forcing men to "sit at home" and care for children. This disruption gave rise to women advocating for labor rights and women's suffrage as they suddenly gained a voice.[6]

Society—people as a whole—greatly benefited. The Luddites were trampled. Should we have held out a hand and cried out for technology to stop for their sake? For their sake, should we have put the women back into the home and left working families with barely enough patched and worn outfits for each member? One might say it's a false choice. The Luddites also didn't need to be violently suppressed. But while they weren't a secretive medieval guild, jealously guarding their craft from outsiders, the widespread benefits came *from* downskilling their trade. Their hard-earned expertise would become useless. Their difficult trade would become something that unskilled women and children could pick up.[7] Clothing became something everyone could have in abundance, not something only the rich had enough of.

In technology transitions, almost everyone wins. But some lose.

APOCALYPSE OF JUNIOR JOBS

Even among programmers, I've always been a militant automator. I've lost track of how many automated scripts run on my home servers. That's fine since they set themselves up with infrastructure tools like Ansible. My various apartments and eventually homes have always been smart long before Alexa, HomeKit, or Matter made it easy. Volunteering for a nonprofit, I obsoleted my data entry position forever by writing the world's only automatic conversion tool (that I know of) for FoxPro, a dead technology even in 2010, to SQL. When I became an editor at one of my college's newspapers, I went on a tool-building spree to forcefully halt as much manual work as I could. I've been known to snatch work from interns and cry out that it can be done programmatically. Those who've worked with me know I regularly say "single source of truth" and "don't repeat yourself" to groups that have no idea what those statements' lineages are.

One of my core philosophies that I've espoused for decades is humans should never be burdened with something a machine can do. The point, really, is that I admit I'm not the most neutral commentator on automation.

But the reason why I've always felt this way is that many tasks people, especially in low-level positions, are sent off to do are *needlessly fruitless*. They are brain-numbing and repetitive because if they weren't, they wouldn't have been able to be programmatically automated before AI.

And that *is* what we've seen over time. Marc Andreessen famously said that "Software is eating the world" in 2011, explaining why new software start-ups were growing so fast and commanding extraordinary valuations. Given where we ended up, he was obviously right. Back then, he decried the lack of education and skills required for most in the US to participate in the software boom.[8] It was an early version of the later tired platitude that people should learn to code.

Today, software engineers vacillate between excitement about "vibe coding" (programmers sparking joy from AI executing on necessary but tedious coding) and pronouncing the death of software development. With the latter is often embedded dark humor about where the mantra of "learn to code" has taken people.[9] Economist and former US Secretary of Treasury Larry Summers even advised against learning to code. AI was going to automate it anyway. Gary Tan, CEO of start-up accelerator Y-Combinator and founder of Initialized Capital, rebutted that Summers' take was "terrible advice."[10]

I tend to agree with Tan. Someone who blindly lets an LLM code for them will eventually be buried under errors and broken code. Remember, the nature of LLMs means that they will *never* reach total perfection—the steps necessary to achieve it would inherently destroy their superpower of generalizability. Hence, you'll always need something to verify them, which we introduced previously—an oracle. In this case, it requires an experienced programmer to be able to quickly see problems in the LLM's approach, correct them, and then repeat. We'll likely get more "automated verification" over time (machines or algorithms as oracles), but we'll probably never totally remove the expert's role. Without it, one is liable to "vibe" one's way into an indecipherable snarl of escalating bugs—especially once a codebase reaches a substantive size with thousands or even millions of lines of code. AI supercharges *expert* productivity.

Take it from me. I've tried most of the AI coding tools—certainly all of the major ones as of this writing. They feel much more like going from a screwdriver to a power drill (or impact driver). It's not like a robot suddenly takes over and builds your IKEA furniture for you. Too much power can also be a bad thing. Hand tools may be slow, but they likely won't strip a screw or crack wood. Powerful tools can do incredible things. Or cause incredible damage.

Beyond that, programming requires quite a bit of creativity. That may be surprising to some non-programmers. I write both code and prose, and while the processes are clearly different, they can often be surprisingly similar. Finding the exact words to capture the essence of something is very similar to writing an elegant, readable solution to a technical problem. A common aphorism in programming (whose true origin is lost to time) is "Code is read much more often than it is written."[11] It's true. In any large production system, collaborators need to read code to understand the context necessary to write their own. Readability and intelligibility are critical. Prose and code are both written for readers *and* are deliberately crafted to serve a specific purpose. Creativity underlies both.

That's why AI can't easily replace experienced engineers. What it *can* do is dramatically reduce the need for low-skill, entry-level programmers—those hired to churn out rote, boilerplate code. Occasionally these roles are outsourced to low-cost labor markets. Often, they're filled by younger, junior engineers who haven't yet acquired the experience and skills to operate on their own. I've managed both. If AI had been as advanced as it is now, we might not have needed junior programmers.

Contrary to wiping out all developers, AI may create a chasm between those with experience and those without. Others have noticed the same. In conversation between experienced developers on Reddit, many noted that—contrary to management's belief that they could all be replaced—they felt supercharged. As one commenter noted, AI seems to act as a multiplier on the code one can already write. Bad programmers with AI make lots of bad code. Great programmers with AI make lots of great code.[12] Again, we understand why. AIs are best with oracles.

A *lot* of fields are like this. Experts bring elements of unbounded variation to their jobs *and* the ability to act as oracles to overcome the statistical, know-nothing nature of AI. Entry-level juniors…bring none of those things. Until they reach senior level, juniors are the executors of tasks that seniors review. The two sectors I have the most personal experience in, technology and finance, tend to work this way. Many, many other sectors operate similarly. With AI, what will juniors—usually younger people newer to the field—do?

As a first-year hedge fund analyst, I spent days doing simple work my manager could do better in an hour. The entry-level programmers I oversaw struggled for hours, writing and debugging code that I could fix in a ten-second glance. Even aside from "apprenticeship-like" roles, many other sectors have rote, tedious work that entrants do to "break into" an industry.

A friend who eventually became a full-time, professional journalist started out barely surviving on odd jobs and writing soul-sapping listicles like, "Top Ten Trends You Need to Know This Summer." My wife's first internship out of college was at a hot fashion start-up, where she did data entry and fetched coffee. A well-regarded translator I spoke with who handles translating complex Chinese legal documents started her career on one of the online gig-work platforms, translating mind-numbing submissions for five or ten dollars.

All of these jobs can be done by AI. Most of these can already largely be done with whatever LLM you want to open in your browser right now. Even my hedge fund job at Bridgewater often had me writing up memos analyzing a market. I know multiple managers running their own funds who have told me they use ChatGPT or Perplexity's Deep Research to do exactly that. It's good at summarization. Terrible at generating ideas. That's basically the same as a junior analyst anyway.

As a junior analyst, I did do things that an LLM can't easily do, though. Whether it was data cleaning, munging, warehousing, highly analytical or quantitative work isn't really an LLM's forte. Most of those tasks are still fairly rote, though, and don't require any of the elements of "unbounded variation" (or imagination). That makes them vulnerable to AI, even if not

LLMs specifically. Those tasks are precisely what a product and company called Rose AI tackles.

Alex Campbell, the founder and CEO of Rose AI, was formerly head of the commodities team at Bridgewater and before that was a prop trader at Lehman Brothers. His platform uses AI to do the kind of numbers and data work a junior analyst would do and an LLM alone would struggle with. I spoke with him in early 2025 about AI in the finance industry.

As he describes it, the reason AI companies haven't broken into the finance sector more is partly technological and partly cultural. On the technological side, it's a fairly specific problem of visualizing, cleaning, and interpreting very freeform data. Modern AI can do it, but it is a specific application. Culturally, technologists—similar to the way they approach medicine—often barrel in and demand an industry do things their way. "We think you should do your work this way… use our structure… move how you work out of this environment," Alex said, mimicking how he'd heard technologists try to sell banks and hedge funds on their technology tools.

In reality, the finance industry is built on Excel. But it also operates with screenshots and emailed files. It's a mess that needs an interoperability layer that Rose AI provides—which, to a technologist's eyes, seems totally nonsensical.

However, as Alex explains, it's a matter of a more regulated industry but also of trust. "When you're in a start-up or tech company, the answer is always yes… 'I can do that.'" In banking or trading, "saying you know the answer when you don't is the worst possible thing you can do… at some point, it becomes fraud." That's partly why you have to use finance's preferred tools, and why finance is so reluctant to change. A bug in a tech company is a small hiccup, a rollback, and relaunching the service. A bug in a hedge fund can be billions of dollars.

That all creates a mighty barrier to simply opening ChatGPT to replace a junior analyst. Except, of course, it was never a permanent one, as Rose AI shows. Even so, similar to what we've already said with writing,

programming, and everything else, Rose AI is not going to automate all of finance.

Alex describes that there used to be coders, spreadsheet people, and analysts—all of these specific, rote roles could be collapsed into "one loop." These tasks are often repetitive and straightforward. As Alex describes it, "You get some winnowing. Right? You need less labor. I made this kind of jokey tweet the other day: 'It's the age of the idea guy.' I think that's really, really the case. That's part of productivity."

Of course, that doesn't mean that it's just a literal "idea guy" that knows nothing about any of those things.

"We have this phrase called 'Don't trust machines,' which leads you to reflect on when the machine breaks," clarifies Alex. He explains that if you ask for US inflation, you can get dozens of versions of it. He describes doing an analysis on silver required for the growing needs of the solar industry. Different estimates can be five times, or ten times, or something else radically different based on your assumptions and what technology is used. "This is the kind of thing you shouldn't really trust robots on," explains Alex. What you *can* do is, once a trusted expert vets the assumptions, ask it to take a model forward ten years. Which is precisely what a junior analyst might be asked to do.

LLMs are the most visible face of AI, but they're just the beginning. Language is only the most visible and obvious domain where AI has tread that previously only humans could go. Inevitably, so long as it isn't barred by "unbounded variation," we will see AI innovate, develop, and move into others—like we see already with Rose AI.

So, does that mean young people are doomed? I remember how difficult it was to get started out of college when I was young. Every entry-level position, inexplicably, would require three or five years of experience. That market was painful enough. What will those just entering the workforce do now?

A sentiment written in a piece titled "AI Won't Kill Junior Devs—But Your Hiring Strategy Might" by Addy Osmani, author and engineering manager

at Google, is that the role of junior developers is evolving, rather than disappearing. He writes:

"Rather than writing boilerplate code (which AI now handles), juniors must focus on higher-level skills like debugging, system design, and effective collaboration. Companies that cut junior positions entirely risk their future talent pipeline. The most successful junior developers will use AI as a learning tool rather than a crutch, verifying its output and understanding the 'why' behind solutions. They should make sure they build up the ability to read and comprehend code."[13]

I think he's right that the role of young talent, in engineering and elsewhere, will evolve. I'm not so sure that risk to a theoretical future talent pipeline is going to dissuade all but the largest organization—to be fair, perhaps like Google. If nothing else, there's a collective action problem. It's better for everyone if juniors get the training and experience they need to become seniors…but *we* don't have to be the ones to bear the cost.

Ultimately, we'll have to see. Maybe it'll be all right. Young people will just have to learn a new technology, as they always have. Or maybe we'll see an extraordinarily difficult transition that leaves many young people stranded without a way to climb a career ladder. I don't think anyone has the answers yet.

WHAT BOUNDED VARIATION REALLY MEANS

One might have come away with the feeling from the last chapter that most jobs will still require humans. So many aspects of our lives require some aspect of "unbounded variation" in creativity, adaptability, or intentionality. Or if you prefer Pearl's formulation, counterfactual thinking—the ability to imagine and think, "What if," without experiencing it. That feeling wouldn't be wrong. Most actually do. Current AI tends to work best *with* humans.

There *are* areas where AI can work without humans, though. Or, even if a few humans are required to oversee the process, it could look similar to modern agriculture, where a tiny fraction of the population is able to do the

job productively enough that they can satisfy the demands of the whole world.

I'm not going to go through every potential job at risk. It would likely be a long one that you wouldn't want to read—but also I doubt I'd be able to come up with an exhaustive list. There are too many corners of the economy and specific individual cases. Instead, the better way I can help is to provide an understanding of what is vulnerable.

Most of these jobs are ones that, frankly, don't add that much individual value. They don't require any significant judgment or creativity and don't vary that much. They just happen to vary *enough* that it's much easier to throw a human at the problem. We've already seen that many of these jobs tend to be what organizations task their junior hires to do. But not all people who do these kinds of jobs are junior or will "grow" out of them.

Some writers will never be a *New York Times* columnist and subsist on articles that are—not to be mean—essentially formulaic clickbait meant to draw advertising or affiliate revenue. Some freelance translators expect to always stick with translating simple, low-stakes documents for a low hourly rate on platforms like Fiverr or Upwork. Some designers will never be featured in glossy coffee table books, charging "luxury" rates, and instead crank out "good enough" wireframes or logos as (one of) the lowest bidders for yet another app. We could go on: photographers who just make stock photos, videographers who make filler footage, musicians who make looping music beds underneath voiceovers, or (speaking of) voice actors whose pieces aren't meant to be particularly memorable but just provide a voice track.

AI is already capable of handling all these tasks. At the end of May 2025, Business Insider announced that they would be laying off about a fifth of their workforce and investing heavily in AI, with the aim of having all employees regularly use Enterprise ChatGPT.[14] Machine translation is "good enough" now that tools exist to make entire websites multilingual with barely any human intervention.[15] Tools like Google's Stitch, v0, Lovable, and bolt are all meant to replace low-level UI/UX (user interface/user experience) designers.[16] We're even seeing property

managers using ChatGPT's image generation capabilities to quickly do mock-ups, replacing their entire predesign process (and people they'd hire to do it).[17] And, of course, we already have AI images, videos, music, and voices—enough that social media platforms are struggling to come up with rules around them.[18]

To be blunt, I have yet to meet anyone who does this kind of work and describes it as deeply personally fulfilling. I don't doubt *someone* probably does, but these tend to be repetitive tasks that aren't particularly mentally stimulating for someone who is already practiced. They also tend to be low-paid, low-stakes, and have a low bar for quality. If they weren't, we wouldn't be seeing this much uptake on AI usage without human oversight.

Why do people do it? It's not always due to the individual's inability to succeed in the "big leagues." Sometimes, there are life circumstances with health or family. Sometimes they simply prefer the freedom and flexibility. For others, they have other priorities in their lives, like a hobby or passion project, and a low-stress, low-mental-strain way to pay the bills is all they want. And while most of these jobs—especially for those freelancing—can never be described as secure, one would always need *someone* to do it.

At least until now. "For the first time in history, technology will take over knowledge jobs," says Makki. He expects what he calls "digital autonomous workers" (DAWs) or "autonomous agents" to be everywhere —"impact[ing] every sector, from education to health care." The term "AI agents" has already become common for AI systems that can autonomously perform tasks, which we have observed penetrating more and more industries.

As he says, we *are* starting to see them everywhere. And Makki would know. As a reminder, he is chief technologist at Mubadala Ventures. Mubadala Ventures and its parent entity, Mubadala, one of the sovereign wealth funds of the United Arab Emirates. They have been one of the most active investors in AI, already putting billions of dollars into the sector.

He elaborates, "DAWs will redefine the nature of work and the work economy… grab[bing] a portion of the global services market valued at

over fifteen trillion dollars. Even if it grabs just five percent, that's tremendous growth for AI." That's also a massive disruption in types of roles that's unprecedented in history.

NOT JUST KNOWLEDGE WORK

Traditionally, manual labor has suffered—and gained—from automation. With AI, knowledge work has ironically been more vulnerable. Using our framework, we can understand why.

Models, data, and compute. Our models and compute are the same as we've become familiar with up to this point, with all of their power and generality. However, only in knowledge work is data readily available. The digital revolution of the past few decades has moved knowledge work onto computers. I'm writing this book on a computer after all, not with paper and pen.

Manual work, in contrast, has not. It's not like construction workers build things on their laptops. Agricultural workers don't bring their computers as tools to pick fruit. Pickers in warehouses are using their hands, not pointing and clicking with a mouse. As I've repeatedly emphasized in my points about data friction, this hinders the ability for AI to enter these areas—but not permanently.

As of this writing, I am already familiar with and have received pitches from multiple companies attempting to develop foundational models for robotics. The idea is to make physical dexterity easily generalized in the same way as language with LLMs.

Counterintuitively, to those outside of the robotics industry, the most significant costs of operating robots—like the giant industrial arms in automotive manufacturing—are not the robots themselves. Often the operations, especially programming (and reprogramming), are expensive. An industry rule of thumb is that the cost of integration, including safety, tooling, and programming, is four to six times the cost of the robot.[19] So, a $100,000 robot could cost $400,000 to $600,000 to actually deploy and use.

Beyond that, it's a surprisingly challenging problem to replicate human dexterity. Picking fruit without crushing it is extraordinarily difficult.

Achieving precise control and effectively incorporating feedback to hold the fruit with just the right amount of force is extremely challenging. It's led multiple start-ups I've encountered attempting to commercialize this technology to ultimately fail. Robotics researchers have joked with me before that we'd likely reach Mars before we get practical, widespread laundry-folding robots.

Data, the lack of it and the friction of physical data, is still a huge challenge. It's also not an insurmountable one. After all, we've already seen one of these area's data and engineering challenges start to fall—self-driving cars.

It's also an area with a lot of potential disruption to jobs. Uber drivers are also freelancers who share many traits with those in "knowledge work."

Self-driving cars aren't theoretical. They aren't even that uncommon anymore. In 2014, I got to ride in a self-driving car while at Google X—which was a unique perk. Now, Waymo, the self-driving car company spun out of Google, is accessible to consumers as easily as an Uber in multiple cities across the US. I have regularly ridden in them in San Francisco. I tried to call a Waymo car in Austin, Texas, literally through the Uber app during the SXSW conference in 2025 (it was the only way to request them at the time). I couldn't get one because the option was too popular and *everyone* seemed to want one.

"Today, I can walk outside my office and get a self-driving car…and it will work exactly like an Uber in many ways. That's a huge testament to everyone who worked in the industry back to the original DARPA Grand Challenge," explained Brandon Basso when I spoke to him.

In 2004, DARPA (Defense Advanced Research Projects Agency) initiated its Grand Challenge as a catalyst for innovation in autonomous vehicles. And Basso? He earned his PhD at UC Berkeley, working on machine learning problems like routing vehicles and drones. That's where we originally met. He was at the Vehicle Dynamics Lab while I was doing my MBA. Afterward, he joined 3D Robotics, an early US drone company, then moved to Uber ATG (its self-driving unit), and finally to Cruise (GM's self-

driving car subsidiary), where he is in charge of simulation and evaluation of their autonomous vehicles.

In his view, the two biggest unlocks have been hardware, sensing technology in particular, and deep learning. With hardware, cameras, radars, and LIDAR (Light Detection and Ranging, which can be thought of as "laser radar"), since the early days of self-driving, systems have become commoditized and can be bought off-the-shelf instead of being custom-made. Basso explains that deep learning significantly advanced computer vision technology. But the bigger impact was moving away from rules-based systems, which simply weren't flexible enough. That was a huge advance.

Counterintuitively, though, it's becoming harder to get meaningful data nowadays. "This interesting thing happens where you drive more and more and you push those systems to their limit. Because they're performing better and better, you see fewer examples," shared Basso. To describe it another way, perfect drives from A to B stop being useful. Within "bounded variation," we want self-driving cars to handle *variation*—exceptions, chaos, and *problems* that inevitably occur in the real world. As we've explained, AI needs data to tackle that variation. Early on, there was a lot of low-hanging fruit (data) to be harvested just to be able to drive at all. Now, as we aim for full autonomy, we need even more scale to see the exceptions. With *fleets* of cars, whether it's Tesla, Waymo, or GM, we got that scale. Machine learning unlocked "the ability to directly connect the data into improvements of the fleet," described Basso.

Hardware, modern AI, and the *scale* of data collection brought the physical data friction of autonomous vehicles crashing down. In reality, there is not an infinite amount of variation—you might say, "unbounded"—in normal driving. Deep learning can handle some variation and generalization. As fleets continue to gather data, eventually every corner case will have been seen. Any others that emerge subsequently, even once, can be incorporated with simulations like what Basso does helping to "bolster" the novel example.

All of these factors have helped make self-driving cars safer than humans on average (though not yet in all conditions).[20] Other than novelty, it's partly why Waymos were so popular in Austin. But some have described to me how *nice* it is to get in a self-driving car, have an extremely smooth ride, and not need to talk to anyone. It's kind of a "reverse" version of the "human touch" factor we talked about, especially for introverts. As a chatterbox who enjoys learning about everything and is traveling constantly, I find Uber drivers from all over the world to be a rich source of information on different jobs, public opinion, and culture. I generally enjoy the experience, so I can't say I feel the same way about conventional human-driven Ubers versus Waymos. Though I do understand. Regardless, consumers are voting with their dollars.

In a report by Bond Capital, the famed Mary Meeker's firm, Meeker revealed that in April 2025, Waymo had taken substantial market share from both Uber and Lyft—*and* now had a larger market share than Lyft. It had 27 percent market share while Lyft had 19 percent. In August 2023, Waymo started with zero.[21]

Self-driving cars aren't alone in changing the physical landscape. A company called Manna, which builds and operates delivery drones, is already delivering coffee from shops to consumers in Ireland with a two-minute and forty-second average flight time end-to-end.[22] Chains like Domino's and Chick-fil-A have experimented on-and-off with autonomous delivery for years.[23] The superpower of modern AI, again, is its generalizability and ability to capture limited variation. With a surprisingly large class of problems in industries like health care or logistics, as long as we have enough data—which we don't for now—it's fairly easy to load it into a neural network and have it work "pretty well."

The changes are here, and they'll keep accelerating. Physical data is a barrier but not an insurmountable one. Once again, only those very fundamentally human aspects of "unbounded variation" cannot be replicated.

AI EVERYTHING

Inevitably, outside of niche cases, everything that *can be* taken over by AI *will be*. That makes me sound like a fanatic, but I have a simple reason for saying this. It's the inevitable math of AI's economics. AI gets cheaper over time. Humans don't. Remember Ark Invest's data that AI costs (inference and training) have been falling by 75 percent per year?[24] *Humans* certainly wouldn't want their labor cost—meaning salary and wages—to fall 75 percent per year.

That's why companies you wouldn't associate with technology, like United Airlines, are pushing it into everything. In mid-2025, United's CEO touted, "We're probably doing more AI than anyone." The company has been experimenting with AI in everything from call centers, sharing flight details, baggage recovery, and predictive maintenance—even if much of it is still experimental.[25]

Google has integrated it into all of its products. As Nathan Lambert at Interconnects points out, citing Google's 2025 developer conference, Google's AI products processed more than 480 *trillion* tokens in April 2025. Lambert contextualizes that number, which puts "the daily token processing of Google as a mix of reading or generating all the data in Google Books in four hours or *all the instant messages stored in the world* in a little over a month."[26]

In a June 2025 blog post, the venture firm Andreessen Horowitz shared that they were seeing AI start-ups grow much faster than past start-ups. They cited examples of companies reaching fifty million dollars in annualized recurring revenue (ARR) in six months—or even one hundred million dollars ARR in a year. "Pre-AI, a common benchmark for best-in-class enterprise start-ups was one million dollars in ARR in its first twelve months," the authors pointed out.[27]

Some use cases are inevitably too expensive right now, even with how much AI costs have fallen. Without giving away confidential information, it *is* possible to build AI coding tools in ways that are so expensive that it's next to impossible to reasonably scale right now. Cobbled-together AI agents comprise these tools, striving to consume codebases spanning millions of lines. That bite is too large for most AIs to chew, even with

RAG support. That bite is also ridiculously expensive. The media often portrays AI as costly and power-hungry because they're covering the frontier. But eventually the frontier will engulf even these kinds of applications, which will then be routine. And cheap.

That's why, again, what *can* be AI, *will*.

Perhaps that's a good thing. The general buzz appears to be that AI may pose a threat to human jobs. Maybe. If nothing else, there'll likely be disruption in the short and medium term. But what about the longer term? AI is emerging during a particularly critical period, where we are finding ourselves short of humans to throw at many of our rapidly growing problems. As much as AI feels threatening now, perhaps in the long term, it's also our salvation.

CHAPTER THIRTEEN

HUMANS + AI

There's something called the "dead internet theory." The idea is that automated bots drive most activity on the web, including on social media. Bots post content. Bots like, comment, and follow it. Essentially, machines talking to machines. With the amount of engagement studies have found from automated accounts aimed at inflating ad revenue or spreading propaganda, it's not totally crazy.[1]

With the emergence of more powerful AI, it's more salient now than ever. We've had social media bots for years. Modern AI has allowed it to bleed into more aspects of life. Just two months after ChatGPT's release, surveys of college students had found that around ninety percent had already used it on assignments.[2] To complete the loop, we are now seeing teachers grading students' assignments using AI.[3] Perhaps it's time to upgrade the "dead internet theory" to the "dead world theory"?

Every technological transition has brought this kind of doomsaying, though.

"If you go back to when the phonograph was first invented, there were demonstrations that this is going to kill education. Because if you can record stuff, you can record all the lectures, and you don't need teachers anymore," Makki chuckled. As an educator himself, a professor in electrical engineering, he's seen his share of technologies come and go in education. But he does expect education should change and wonders if the system is up to it. "If the global economy changes, education should change," he relates.

That change isn't just education becoming AI talking to AI any more than our system evolving from classrooms to students sitting at home listening to phonograph recordings. We *do* have online courses and platforms now. I

was in one of the first classes of OMSCS (Online Masters of Science in Computer Science) offered by the Georgia Institute of Technology. It was the first of its kind to grant a fully offline equivalent advanced degree to online students. That program was facilitated by technology, which allowed it to scale beyond what an on-campus one could, but it was still extremely hands-on from faculty and graduate student instructors. Having taken graduate-level computer science seminars in person at UC Berkeley, I wouldn't say the rigor suffered either.

Technological tools change our lives. Shovels dig better than bare hands. Calculators calculate faster and more accurately than people. The internet, audio codecs, and speakers carry one's voice much further than the loudest human shout ever could. None of these have obsoleted humans.

In a world where AI can and will do many tasks—including those we never thought possible by a machine—it's helpful to keep that in mind. It's still a tool. For all of its power, it has no motive, will, or intentionality of its own. Despite how human-like LLMs can seem sometimes, they really are just statistical models—tools.

Now, tools may cause fewer humans to be required to do a certain task. I'd wager that with modern heavy machinery, one could build the pyramids of Egypt with less labor than it took the pharaohs thousands of years ago. So, as we've rather extensively covered in the last chapter, it isn't to minimize the kind of disruption they'll cause, which is potentially quite significant. But at the end of the day, tools won't drive themselves. They work *with* humans… the supply of which is diminishing, relative to what we need.

A WORLD WITH FEWER HUMANS

I've sometimes wondered what AI would do in a world where we had *too* many people. A world where our population was rapidly expanding, and we couldn't find enough jobs to keep our many hands busy—or fed. This world would see the cost of labor, and hence real wages, collapsing and the multitude becoming worse off as human labor just isn't worth very much. There's too much of it.

It’s hard to say if that world is even possible. It resembles the world described by Thomas Malthus in our first chapter. Economic theory, and history, would suggest that we’d figure it out. Admittedly, the “figuring it out” was largely done on the back of technological advances—making human labor much more productive and hence more valuable.

Today, we’re faced with a rather peculiar challenge. It’s the opposite of too many people. Our population growth is decelerating. Our population is rapidly aging. And, as a result, we’ll have far less labor than we need to maintain our current standards of living—let alone improve them.

Soon, if not already, the world will begin losing population. Many high-income countries have been below replacement rate since the 1970s.[4] With better access to contraceptives and improving conditions, the developing world—even sub-Saharan Africa—has seen its fertility rate starkly plummet.[5] Intuitively, it takes two children per couple to replace the couple. Adding in accidents, disease, and other deaths, you need something above two. Researcher Jesús Fernández-Villaverde estimated that global fertility fell below 2.2 in 2023, breaching his estimate of the world’s replacement rate of 2.22.[6]

As a whole, this is a good thing. More planning and choice for parents. And far greater opportunities for economic advancement for women, who have always shouldered a greater burden from childrearing.

Unfortunately, it’s also an economic problem. Neil Howe, a demographer at Hedgeye Risk Management, notes that worsening demographics could cause a lost decade in economic growth.[7] I agree with Howe. Why? The core of the problem is, to put it bluntly, older people who leave the workforce still consume goods and services. They still eat food. They still live in houses. They still need various things that everyone else needs, and wants, in life. However, they can no longer put in labor to produce it. This means, effectively, older individuals in an economy become dependents.

One concept that I have watched very closely since my hedge fund days at Bridgewater has been the dependency ratio of economies. Conceptually, it’s the ratio of older adults (retired or about to retire) relative to the working

age population that can produce economic goods to support them. More precisely, it is the number of over sixty-four-year-olds divided by the number of fifteen- to sixty-four-year-olds.

As we can see below from World Bank data I retrieved from Federal Reserve Economic Data (FRED), since 2000, that ratio has been skyrocketing since 2000.[8] The whole world has been aging.

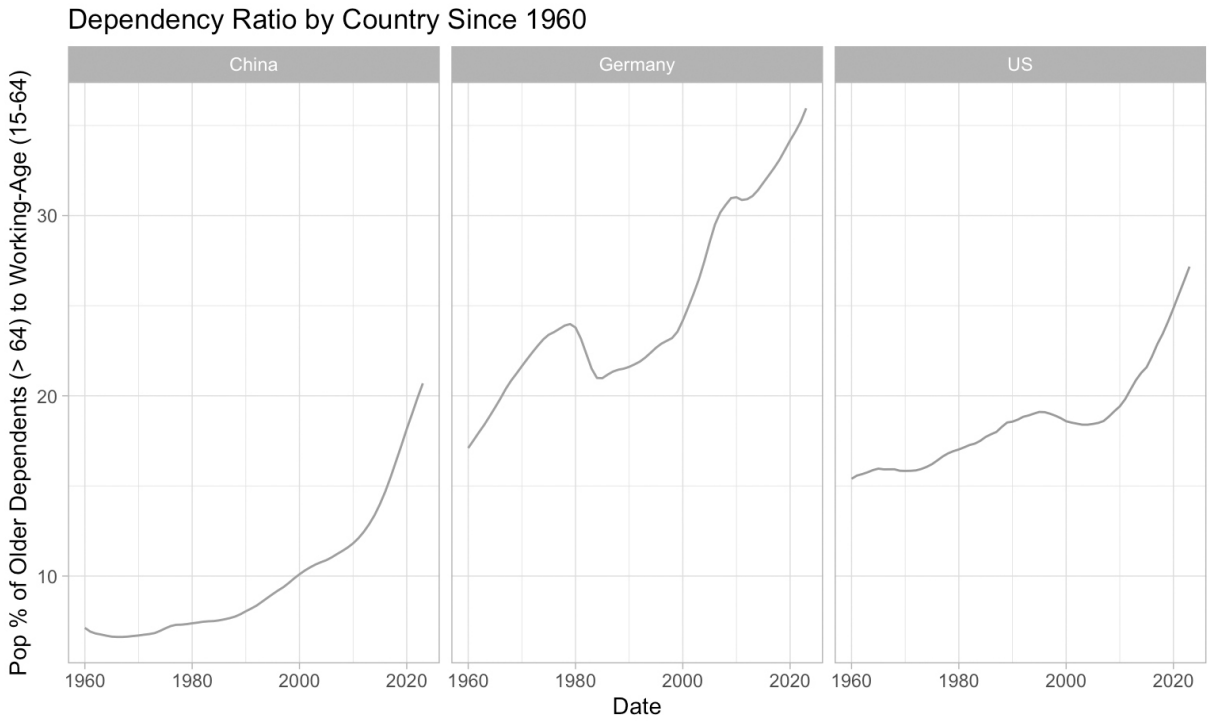


Chart by author using FRED data. Dependency ratios for China, Germany, and the United States. While each has their own unique circumstances, each representative of Asia, Europe, and North America similarly sees massively aging populations since 2000.

Attempts to reverse this trend have largely failed around the world. Japan, which has confronted the problem the longest, has attempted a multitude of plans, including parental leave and subsidized childcare. Nothing worked.[9]

If nothing is done, we will see all sorts of problematic impacts. South Korea, another country struggling with low fertility, has one of the world's largest pension funds and is on track to be depleted and unable to pay pensioners by 2056.[10] There are too few working payers and too many

pensioners. Every few weeks, I encountered similar dire, underfunded (meaning too little money in the pension fund relative to upcoming payments and expected investment returns) pension funds either in my analysis or speaking with investment staff at these institutions. However, pension funds are merely the most visible and easily quantifiable impact.

Macroeconomics is notoriously difficult to intuit. However, I will attempt it anyway. Fewer working people producing things (whether goods or services) and more people consuming them literally means a lower GDP (gross domestic product) per capita, one of the most universally accepted measures of wealth and prosperity.[11] Less stuff, more people. Simple enough. Translated from any technical definition, everyone gets poorer.

For many of my readers who have never lived in economies experiencing true economic contraction, not merely cyclical recessions, the reaction may be, “Oh, well.”

No. It’s nearly impossible to overstate how much people *hate* getting poorer.

A review of economic and conflict data by the *Journal of Conflict Resolution* notes that while conventional wisdom connects economic recessions and coups, data is sometimes mixed. By removing more transient shocks and taking into account only more persistent contractions, the connection between coups (i.e., violence) and economic contraction is remarkably robust.[12] The IMF, in a comprehensive study, finds a similar result between economic downturns and conflict, going even further and quantifying greater fatalities with greater shocks.[13]

In the US, the 2008 financial crisis rapidly brought thousands out to the street in protest—and arguably infused US politics with much of its later tumultuous energy.[14] Inflation, which also causes regular people to feel poorer, came roaring back into the world after the pandemic in 2020. After a few years allowed people to truly *feel* its effects, political incumbents lost all around the world in 2024. We also saw extreme political movements emerge with citizens demanding radical change *now*. As one article put it,

"Inflation is poison for incumbents."[15] These recent examples are relatively *mild* compared to what we could experience in the coming years.

Economic pain isn't, "Oh, well." It causes unrest, anger, and pain. And in extreme cases, it creates violence. People *die*. And in much of the world, that's horrifyingly literal.

So what can we do? As discussed, attempts to raise the fertility rate with various incentive plans have failed. And I'd personally rule out the "solution" of reversing gains in women's rights, family planning, and other measures that improved the material well-being and autonomy of women, which were all correlated with drops in fertility. We *could* go with the other direction of decreasing unproductive mouths to feed and clothe, like the mythologized "death panels" of the 2009 Affordable Health Care debates and "pull the plug on grandma." Not only does that seem immoral, but it's also (putting aside the actual veracity of the term during the US health care debates) unpopular.[16]

So, we can't increase the number of young people, and we can't decrease the number of old people. It's a simple equation. That leads to lower standards of living and all of the pain we discussed. Or does it? What if each working person was *massively more productive*, enough to overcome the drag of a larger dependent population?

Now, what kind of technology could *possibly* effect such a change?

JOBS THAT NEVER EXISTED AT ALL

Inevitably, jobs that get erased will get all of the headlines. However, there's a vast class of jobs that no human would be doing anyway. I'm not merely talking about the rather overused example of developed country residents not wanting to work in fields or factories—though certainly cases like this exist.

For example, one of my invested companies is a pizza-making "robot" piloted at Domino's and serving pizzas from multiple stadiums and universities. Admittedly, it looks more like a conveyor belt that adds ingredients using 3D printing technology and computer vision than something from the Jetsons. That's what makes it cost-effective, but that's

beside the point. Speaking with the CEO at the time, he described the robot as “taking away job openings, not jobs.” Many shops could barely keep their locations staffed, which meant a robot didn’t eliminate the, often last, employee. After all, who was going to run the machine if there were zero? Instead, it made that employee’s life less crazy, which improved production and service, which also meant the locations could be more profitable and pay the employee more.

The job used to require more people, though, before our demographic crisis. The robot and AI may be eliminating a job opening versus an actual job, but it was still previously a job opening for a human. What about jobs that no one was *ever* going to hire for?

I spoke with one personal trainer who drives for Uber part-time, which is how I had a conversation with him on a long drive between Boulder and Denver. Instead of manually typing up his notes and training logs, he uploads iPhone pictures of his paper records to ChatGPT and asks for spreadsheets. He then shares those with his clients.

What did he do before? Well, nothing. In theory, he could have typed it up, but it was never worth his time and not valuable enough for most of his clients. He could have hired someone to do the data entry—but that was also never going to happen. Even if it was five or ten dollars per session’s logs, he would never be able to justify it. Instead, he used to just give high-level summaries of progress or simple texts like, “Good job!” Some of his clients have found it extremely useful to have the detailed data. But, he figured, none of them would have been willing to pay a significant upcharge before ChatGPT made it easy for him to do it himself.

Similarly, I know multiple writers on Substack who regularly use ChatGPT, Midjourney, or other AI image generators to make thumbnails and social sharing images. What did they do before? Well, they either just used a random image or chart from the article, found a random stock photo, or (at most) found a meme template where they plugged in something related to the article. In theory, they could have hired an artist to do the images, like some of the largest social media influencers, but that was never going to happen at their size and level of earnings.

Other cases are where a job that could only be justifiable for a huge corporation can suddenly be brought down to an individual level. I spoke with Filippo Marino, founder and CEO of Safe-esteem, who created the industry's first enterprise risk intelligence program—Global Risk Intelligence and Executive Protection—for McDonald's. There, he managed five analysts for their program. Headcount was obviously the majority of the cost. And while it *would* be nice for a normal person to have this kind of personalized, up-to-date analysis and risk assessment, few are going to actually be able to pay for it.

Marino and his cofounder—who ran a similar program at Coca-Cola—rolled out Safe-xplore. At first, they provided a travel risk app mainly targeted to security and executive protection professionals (i.e., bodyguards), but they have now rolled out a version for personal use. The biggest cost, after all, used to be analysts doing the grunt work. Now, much of that can be automated, with high-level analysts—like Marino, his cofounder, and those they hire—acting as oracles.

None of these jobs existed. They could have *theoretically* existed, but no one was actually going to spend the money and time. AI, especially with rapidly decreasing compute and power costs, *created* the opportunity and generated value where none existed before.

There will be many pockets of opportunity like this all across the economy. Many may never be counted or noticed at all—at least in most traditional measures of GDP. AI company earnings will reflect some of it, but it's not like the personal trainer's total additional economic output is counted anywhere. But it's still true value that's generated for his clients. *Society* captures that value generated by AI, which is a very good thing.

Finally, there are classes of jobs that humans are aging out of—and not training up enough young people to take their place. It's understandable we'd see this phenomenon, given the demographic challenges we're facing.

Welding is one that I've seen a few robotics companies go after because not enough skilled individuals are left in many developed markets relative to demand. Electrical and electronics engineering is another—which is ironic,

given our increasing demand for circuitry and digital “smarts” in everything.

CRITICAL SHORTAGES

Electrical and electronics engineers tend to be a grumpy lot. It isn’t universal, but in my experience, it’s *almost* universal. It is understandable, though. Their jobs are difficult and getting more so with the complexity of modern electronics rising. And they tend to be underappreciated, especially relative to their software colleagues.

“If you’re an electronics engineer from Stanford or whoever going into the workforce, you can be hired by Google, some software company, or McKinsey that’s going to pay you triple the salary on entry. So there’s no point in you going,” laughs Tomide Adesanmi, cofounder and CEO of Circuit Mind. He’s one of the exceptions to grumpiness. In late 2024, I interviewed him about the state of the industry and his company. Once again, for transparency, Circuit Mind is also a company that Creative Ventures has invested in.

“So, actually, the industry is trying to train apprentices. But it’s the same thing. The apprentices spend seven years in the company from the age of fifteen. When they’re twenty-two, they can probably get a much better job in software or something… so they’re hemorrhaging people,” expands Adesanmi.

Circuit Mind generates electronic designs for circuit boards—those, typically green slabs, that have traces of copper and dotted chips all over them. These boards exist inside computers, phones, cars…and almost literally everything these days.

Adesanmi used to build helmet-mounted displays for fighter jet pilots at BAE Systems as an electronic systems engineer. While there, he found out quickly how painful electrical design really was, even for fairly basic systems—let alone ones as complex as those that go into a fighter jet.

The typical process goes something like this. An engineer would create a block diagram of the design, a high-level functional “sketch” of the overall system. Then, the engineer needs to select components. The process

requires them to read complex technical documents called data sheets. These data sheets can be tens or even hundreds of pages of dense technical information and charts just for a single tiny part on a circuit board. This process is painful, laborious, and requires a lot of work just to select the right components, connect them together, analyze, make sure it's all correct and actually fits all of the engineer's requirements.

But that's not the end.

After completing all those steps, you need to fit these components into an extremely tight space—the physical board. In consumer electronics like Apple products, the allotted space only gets smaller and smaller. Typically, Adesanmi relates, this process takes between a week and six weeks per *iteration*. There can be between three and ten iterations to get to the final design if things go well.

Circuit Mind uses AI to bring that process from one to six weeks to somewhere between sixty and six hundred seconds.

One side of the system, called Elixir, uses LLMs to ingest data sheets and pull out relevant information for the engineer in a standardized, user-friendly format. Data sheets, decidedly, don't come like that out of the box. It also isn't just dumping it into ChatGPT either. They use RAG (Retrieval-Augmented-Generation, which we touched on in chapter six) to provide LLMs with relevant context at the right time and significant other custom deep learning tools that plug into the process. The infusion of all of these tools adds semantics and meaning—extracting the information necessary for an engineer to make informed decisions. Such an approach turns a highly analog data sheet into something almost like "digital" data. The engineer can then use this relevant information to easily select components, set constraints, and specify requirements for the design.

The other side of the platform takes the standardized model and uses deterministic, mathematical algorithms to "solve" for the circuit after all design information is populated. Harkening back, this looks much more like a symbolic system than a "connectionist" system like deep learning.

It’s obvious how impactful this system would be in all sorts of industries and would completely revolutionize the process of electronic design. Given how significant this is, one would think, this *must* impact some jobs, right? Again, Adesanmi: “It’s not going to be replacing people because there’s already a massive shortage of engineers in the industry… something like 50 percent of people doing this job are going to retire in the next twenty-five years or something.”

Compounding the problem, the demand for complex designs is skyrocketing. “It’s increasing over time because of [Internet of Things], robotics, machine intelligence, and video. It’s about building lots of GPUs and the new generation of [electronics],” says Adesanmi. “So it’s just exploding. Right? The demand for electronics. So [we are] nowhere near replacing the engineers… because there’s a massive gap in demand and supply right now.”

Experienced engineers also do more than just painful rote work. “These things are just tools that augment an engineer,” points out Adesanmi.

“It’s not something that is going to look like a human engineer. The analogy I make is it will transform what the job means, what an electronic engineer [does]. So you want the more creative things to be done by the engineer—the new things, the new designs, the verification and checking of a design.

“Creativity, that’s one thing that will not be automated. The ownership of the design… making sure it’s right… The platform will help you generate ten different design options and optimize for cost, size, power, change the processor, change this or that. So you’re running through the design space… with a very powerful tool, but the final decision is made by you. And there’s a lot of nuance behind the choice… which can’t be rules-based.”

Once again, we see creativity and adaptability as a key part of what a human brings to the equation. But beyond that, Adesanmi predicts his will allow engineers to be more generalist and high-level. This echoes what we heard from Campbell and Rose AI related to financial analysts.

But while there *are* specialist areas like analog and RF (radio frequency) design, this generalization is *already* de facto happening with fewer and fewer electrical and electronics engineers available. AI won't be disrupting humans here; it will be saving us from the disruption that is already happening in demographic trends.

OUR FUTURE CHALLENGES

We have a lot of problems. Unfavorable demographics is a big one. I spent a lot of time on it in part because it's often poorly covered in most media. A slow-moving, decades-long train wreck stops being newsworthy after a while, after all. The other reason is because it's particularly relevant to the question of jobs and AI. It certainly isn't our *only* problem.

There's climate change, high global debt levels, political polarization, and rising global tensions—just to name a few of the challenges that we already face. People feel it too. For the first time since the 1900s, religiosity among Americans has abruptly stopped falling. Gen Z is flocking to religion—more orthodox versions of Christianity like traditional Catholicism and branches of Calvinism especially. While there are many potential explanations, one twenty-seven-year-old drawn to an orthodox church explains her own reasoning: "We're in a world that is constantly broken. By the time I was twenty-seven, like seventeen major world events had happened in my lifetime. We have to find something concrete, real, and honest."[17]

Maybe AI will make these things worse. With climate change, even with rapidly falling power costs for the same model, we keep pushing the limits of AI—to power-hungry extremes. Perhaps countries will spend themselves into more debt to pursue an arms race in AI. We could see AI flood the internet with even more sophisticated propaganda and misinformation. And maybe the threat of other nation-states getting more powerful AI—or even AGI—first could prompt dangerous escalation.

Or, hopefully, AI can untie the Gordian knot of our aging and falling population gracefully. Instead, it could advance materials engineering and other fields to help solve our climate challenges. Perhaps rising prosperity

from an AI boom could ease our debts and ease our societal—and global—tensions.

Who knows? The truth may be a mix of scenarios. It likely won't be one-sided, though. I'm optimistic in part because many of these challenges don't seem to come with easy solutions as it stands. A Hail Mary with AI could be just the thing we needed.

Exactly how it'll turn out is not something anyone knows. How AI will turn out is something yet to be determined. I don't have a crystal ball, after all. Even so, as we approach the end of this book, I'll try to peer into one anyway.

CONCLUSION

PEERING INTO THE CRYSTAL BALL

Predicting the future is hard. For all of their weaknesses, like their inability to be creative or imagine “What if,” machines are still better than humans at forecasting complex, irregular, but ultimately predictable events like the weather, disease outbreaks, and accidents. As Marino shares regarding the concept of risk, “All of our evolution has left us, without question, with a very sophisticated threat and hazard response system… It’s really sophisticated and deeply rooted, but it’s solely tuned for your present self.”

Of course, this is why, he explains, we built barometers and forecasting models that do better than we do. “Severe weather has been the survival challenge for humanity since the beginning of our days,” he states. We often forget about it today—especially in places like California—but it was a true danger to us back then. Realistically, it still is, even as we’ve built tools to predict the paths of hurricanes or to be able to provide tornado warnings to mitigate its threat to human life.

We created a partnership with our tools and our machines to bring us much further than we could go alone.

Steve Jobs quoted this very idea in an interview in the early days of Apple:[1]

“I remember reading an article when I was about twelve years old. I think it might have been in *Scientific American* where they measured the efficiency of locomotion of all these species… Humans came in about a third of the way down the list, which was not such a great showing for the crown of creation. But somebody there had the imagination to test the efficiency of a human riding a bicycle… [and was] all the way at the top of the list.

“It really made a big impression on me that we humans are tool builders and we can fashion tools that amplify these inherent abilities we have to spectacular magnitudes. And so for me, the computer has always been a bicycle of the mind, something that takes us far beyond… I think we’re just at the early stages of this tool… I think that’s nothing compared to what’s coming in the next hundred years.”

The original Doom was released in 1993. It was revolutionary at the time with clever mathematical tricks that made it possible to simulate 3D graphics.[2] By today’s standards, it looks like a pixelated mess. The other day, a friend showed me a video of the big-budget, AAA game Cyberpunk 2077—with an installed modification to boost the graphics even more called DreamPunk. For many of the shots, I couldn’t quite tell if they were video game graphics or a real person cosplaying as characters from the game. Nonhuman objects like cars, of course, were basically flawless. Humans are still incredible at recognizing other humans, but modern graphics are getting close to fooling us.[3] Those GPUs that helped get us here are now running AI.

SPECULATING ABOUT THE FUTURE

Imagining what will happen next for AI is exactly what AI is bad at, which means we’re on our own here. I wrote this book with the fast-moving nature of the field in mind. Most of it covers principles and history that were meant to be timeless and universal… for a few years anyway. One example is reasoning models. I originally explained their principles and speculated that they’d likely be used more and more. Falling inference cost per token and the ability to scale on a new dimension—more tokens for better results—pointed in that direction.

By the time we entered a later phase of editing, I had entirely removed the speculation. It was already here and used by every major model maker. The outlined concepts were all still correct, but a brand-new book that still spoke about those models gaining traction as a hypothetical felt like releasing a new science textbook that has the sentence, “One day, humans will walk on the moon.”

That’s still going to happen. It’s inevitable. Readers are going to have to live with that. And most importantly for this book ever releasing, I’m going to have to live with that. As a compromise (to myself), readers will find a QR code for a website at the back of this book providing notable updates, errata, and simply other resources to explore. All that being said, it is likely a worthwhile exercise to continue to put together the pieces of what we learned and extrapolate forward. Even if some of this will seem quite outdated in a not-so-distant future.

One aspect is the nature of data. High-friction data, especially that which touches the physical world, will still provide a moat for start-ups. It will also slow the diffusion of AI in those sectors. LLMs, which will likely continue to capture headlines given their greater accessibility to the broader public, will continue to evolve in how they access data. I’ve needled them on this point throughout, but bluntly, the companies making these models have largely just taken it. And if it’s not free to be taken, by definition, it’s stolen.

A former Meta executive and UK deputy prime minister, Nick Clegg, recently was widely reported as saying that AI is doomed in the UK if artists want tech firms to ask for permission to train on their work. That framing *is* flattening a somewhat nuanced point Clegg was making, but it does reflect broad sentiment—both among artists and tech companies.[4]

Going forward, my speculation is that this behavior will diminish. Similar to the early days of digital music on the internet with Napster and file sharing, everything was pirated back then. iTunes and later music streaming platforms changed that. As consumers were able to easily access reasonably priced, legal alternatives, widespread piracy faded. It’s still the Wild West right now in content licensing as *data*, not end-user content. It’s a mismatch like traditional radio and CD business models trying to (not) adapt to digital music early on. I expect a similar path ahead for models and content owners for data. The *New York Times*, after all, just inked a deal with Amazon licensing its data for AI training—the same publication that sued OpenAI in 2023 for using its content for ChatGPT.[5]

Why? Well, the original data is still stored somewhere. The base models are now trained and work pretty well. Most LLMs now also search the web directly using tools and feed in results using RAG to provide the most relevant context for "timely" results. More random text is similar to more uneventful drives for a self-driving car, as Basso described it. It's already well represented and likely not useful. Instead, specialized, niche, or unique datasets will gain value and likely have negotiated licensing rather than straight-up piracy.

This also means that worries about "data poisoning" with LLMs generating more data from the internet—which is then used to train LLMs—are overblown. That argument is roughly that this recursive loop would eventually turn the models into slop, or at least prevent them from going back to the well. Realistically, we've already moved beyond that with sophisticated techniques in post-training with fine-tuning, RLHF (reinforcement learning human feedback), and simulated data. The point is less to know what most of those terms mean. The point is that we aren't just blindly dumping in more data anymore. Additional data fed in will likely become more curated.[6]

Speaking of LLMs, they will likely continue to evolve. While new deep learning models—"normal" neural networks or, more likely, hybridized with other techniques—will continue to gain prominence, LLMs serve as an easy interface to them for humans. Even if what's *really* underneath is no longer an LLM, which is already starting to be the case with more and more tool use like web search or calculations, it'll be what consumers see. As Ben Thompson from *Stratechery* keeps saying and even OpenAI CEO Sam Altman admits, ChatGPT is a consumer brand now with incredible name recognition.[7] Their ultimate success and differentiation may no longer come down to their models as much as "ChatGPT" is to "LLM" as "Kleenex" is to "tissue paper."

Which is ironic because Google should have been in that position. Even so, while OpenAI burns money, Google continues to invest while being an extraordinarily profitable company. And Google is still no slouch. If current trends continue, by the time you read this book, Google AI will be processing more than one quadrillion tokens per month.[8] One writer

estimated the words spoken in a person's lifetime as an oddly specific 860,341,500.[9] Assuming that's true—and given that a token is roughly 0.75 words—that means that Google's AI alone will soon process more words than a person ever utters in a lifetime every three seconds.[10]

We will see whether or not the cost of training new models will start to hit the levels required for a few of these players to become a natural monopoly—and avoid being commoditized. That being said, the industry is still in flux, including with global players, especially China, also competing in AI. But maybe neither of them will "win," and the entire industry will instead move onto open-source models (or, perhaps more accurately, "open-weight models," since the source code and data for training are often not released). Meta's Llama, Mistral's models, DeepSeek's V3/R1, and Alibaba's QWEN number among these open-weight models. These models provide the ability to customize and fine-tune the weights—since they're accessible—for various use cases.[11] They also have less lock-in, since you can always run it on your own machines. If this becomes common enough, OpenAI and Google may eventually also move their cutting-edge models (or some version of them) into the open-weight column.

Nevertheless, whichever way that competition goes, one thing is fairly clear. With volumes skyrocketing and costs plummeting, we're going to have more and more AI services.

Still, this won't change the basic nature of AI. We might get more symbolic AI mixed into "connectionist" deep learning AI (or, as many LLMs now do, push calculations to literal calculators). Use the right tool for the right job, after all. Generally speaking, however, these highly flexible systems will still be imperfect. That means human experts—oracles—will still be critical in verifying and maximizing their value.

Additionally, we aren't likely going to overcome the "unbounded variation" or "counterfactual thinking" barrier anytime soon. As we discussed with Morffis, we can speculate on *how* we could do this, but we haven't done it. Makki shared that he doesn't expect to see AGI, humanlike general intelligence, any time soon—at least if you don't water down its definition.

That obviously still leaves plenty of room for humans to be more productive than ever. We need, after all, to face the many challenges that we'll forced to overcome in the coming decades, not the least of which being our demographic crisis.

Still, I do wonder how entry-level jobs will evolve. And going even further, I wonder whether we'll see even worse income disparities from AI. The internet and globalization already did some of this. Suddenly, everyone could compete everywhere, which meant the best of the best could scale. And those who weren't the best often lost business, given they were no longer protected by geographic barriers, getting by as the only provider in town. This happened to the media industry, e-commerce, and more.

AI has a similar concentrating effect. The translator I spoke with admitted that she had more business than ever. Her nuanced expertise in translating precise legal documents and difficult literature meant she was sought after and highly paid. But she was now able to take on *far* more work by using AI to translate the base text and simply manually reviewing the parts she knows are trickier. Meanwhile, the Society of Authors had shared in 2024 that a third of translators were already losing work to AI.[12]

"Techno-trillionaires" is a term I've heard thrown around in San Francisco, usually fearing the wide disparities created by AI CEOs like Sam Altman and Elon Musk accumulating more power with their model companies. Whether or not that'll actually happen, the far more pervasive aspect of massively enhancing the best in their fields with what could virtually be called superpowers is bound to have profound impacts on society.

That being said, we *do* have a shortage of people relative to our needs. Historically, we've also never run short of the ability to create new jobs in the face of technological disruption. While it may be hard to predict the future, especially with fundamental change or exponential progress, it is Malthusian to simply expect that it's not going to work out.

Ideally, governments will do what they're historically meant to do—help ease these transitions, provide essential services, and regulate in a way that

balances technological innovation and disruption. Whether they do or not is a different story, but that's beyond the scope of this book.

WE'RE NOT DONE YET

Periodically, media commentators ask whether or not AI has hit its ceiling. In practical terms, we are still likely quite far from reaching that point.

Don't get me wrong—we *will eventually* have a "top of the bubble" with a burst. Frankly, that's always happened with everything from railroads at the turn of the century to the DotCom boom and bust. Both still radically changed the world, even after their respective collapses. It's just hard to know what the "top of the bubble" is.

In a way, it is the most rational thing in the world to overestimate a new technology. If you underestimate it, you risk being left behind. Besides that, constant chatter about the "top of the bubble," especially while the technology is still outrunning expectations, creates a dynamic where voices cautioning against exuberance lose credibility. So, yes, we will likely eventually run into a "bubble burst." But similar to the aftermath of the DotCom bust, companies like Google and Amazon didn't go away. They instead became some of the largest companies in the world. AI isn't going to go away either. It's also going to keep evolving.

In May 2025, Google published a paper on AlphaEvolve, which aims to use an evolutionary model—with LLMs throwing out ideas and with automated verifiers (in my terminology, oracles, though these are not human). So, literal evolution. Basically, it acts as a coding agent that can use principles similar to natural selection to refine toward better answers. As Devansh from Iqidis described it in his newsletter, evolutionary algorithms are a mainstay of classical AI. It's another instance of crossovers creating powerful "hybrids."[13] Not that most in the industry are really keeping score of which school different techniques come from versus simply seeing what works.

I could keep going. For instance, I could talk about Google's Titans architecture, which incorporates aspects of memory.[14] This addresses one criticism from Morffis on how LLMs fall short of AGI. I could also speak

about the renewed interest in recurrent neural networks, which share similarities with transformer-based models but come with different tradeoffs. They can scale to longer context windows—the length of the conversation before LLMs start forgetting—hence, they can allow for more powerful *inference*. But they are far worse for *training* because they can't be parallelized for that as easily, making them less scalable in practice versus transformers. Some recent research is explicitly tackling this limitation.[15]

Developments will keep coming, mostly quietly and unseen by the general public until they get rolled out—packaged or accessed with one of the major LLMs. Make no mistake, however: We are continuing to progress rapidly.

THRIVING IN AN AI WORLD

AI, as a major breakthrough and Bob Work's Third Offset, is both scarier and less scary than the so-called First Offset, nuclear weapons.[16] On one hand, there's no risk of us all dying in nuclear fire—unless you take the *Terminator* series literally. And we already discussed why you shouldn't.

On the other hand, AI is going to assert itself far more into our daily lives. While the potential for nuclear apocalypse caused schoolchildren to need to do "duck and cover" drills periodically during the Cold War, AI will constantly surround us, likely embedded into every computing device we own. Also, the bomb never threatened to take your job.

In the midst of geopolitical competition for semiconductors and advantages in AI, individuals can understandably feel helpless against the massive forces—and trillions of dollars—at play within AI. As we've discussed, it is indeed going to disrupt things. It will also create opportunities.

Bill Gurley, a venture capitalist and partner at Benchmark, wrote a now-famous rebuttal in 2014 to an NYU finance professor's article on FiveThirtyEight, "Uber Isn't Worth $17 Billion." The professor's best estimate was that Uber was worth $5.9 billion. The professor's methodology was very standard financial modeling. He used the total available market (TAM)—based on global taxis and car services—and

multiplied that by an assumed market share (10 percent) of the TAM. As an investor and board member of Uber, Gurley pointed out all the reasons he thought the analysis was extremely off.[17]

Between radically faster pickup times, higher coverage, easier payment, and better trust and safety (with driver ratings and Uber's corporate backing), he thought the market would be *much* larger than what the professor estimated. It's a massive market *expansion* with use cases that people never imagined using a hired car for. He talked about an alternative to rental cars, couples' nights out, or transporting kids or older parents. He also thought a great company would be able to get far above 10 percent. Gurley thus proposed that he thought Uber could be easily worth twenty-five times the professor's estimate—around $150 billion.[18]

As of June 2, 2025, Uber is worth around $175 billion on the public market.[19] With food delivery, package drop-off, and many other services, Uber also went much further than even Gurley's imagination in his original post.

Gurley remained fairly civil throughout, but it's notable that he couldn't help but point out that specific section from the professor's original blog post was left out of the FiveThirtyEight version:

"As I attempt to attach a value to Uber, I have to confess that I just downloaded the app and have not used it yet. I spend most of my of [sic] life either in the suburbs, where I can go for days without seeing a taxi, or in New York City, where I find that the subways are a vastly more time-efficient, cheaper and often safer mode of transportation than taxis."[20]

With any new technology, it's important to get your own feet wet. Random commentary, even from supposed experts (who in reality often aren't), will rarely serve you well. I will also say—self-servingly perhaps—that your seeking out this book to learn more about AI is *already* a powerful step to thriving in a world full of AI.

I have a fairly privileged position as an investor and stakeholder in multiple AI companies—even if they aren't ones that are known to everyday people. It's an easy thing for me to give the oversimplified advice of seeking the

areas where “unbounded variation” and imagination are required. Become someone who offers something unique with those requirements and maximize the potential of AI to increase your productivity. If you’re in a role that does not have those aspects that can’t be replaced by AI, look for your exit now.

Don’t get me wrong. Those things are true. But it also leaves out aspects of personal interests, circumstances, and ability.

If you can drop everything and orient your entire life to prepare for the wave of AI sweeping over everything, great. But many can’t just do that. In those cases, it’s a better idea to keep being curious and learning about AI because it’s not suddenly going to stop. Use this book as a starting point to then be able to contextualize commentary, newsletters, and even technical papers. And then be ready for your own opportunity.

It may be a better way of doing things anyway. One of the concepts business school hammered into me—and was actually correct—is the idea of serendipity. You don’t build a network to “get something now.” First off, ideally you like people and don’t just want to use them. But even if you don’t, and even if there isn’t a “match” of your needs with what they can help with, there will likely be something someday. It’s a utilitarian reason to meet people and help others—“karmically” putting positive energy into the world.

In start-up investing, I’ve always done the research ahead of time before seeing the “right company” to invest in. That’s how I understand what the “right company” is to begin with. However, actually *finding* the companies requires putting myself out there and giving serendipity an opportunity to work. If all you do is look at what you have in front of you—either start-ups or personal opportunities—it’s easy to merely rearrange pieces rather than truly imagine something new. You fall into the same fallacy as the NYU professor who couldn’t imagine Uber could be more than the taxi and car service industry as it already existed.[21]

The internet also took a while to take the form that it has today. In the midst of it, it went through blogs, “everything dot com,” and then eventually

stabilized a bit more onto social media, entertainment media, and e-commerce. It also allowed many people to find their own niches, whether with their own unique arts and crafts or sharing their unique nerdiness on a specific subject—now possible because anyone in the world can follow them.

AI will disrupt. AI will create. Because all technology always works that way. I think it's reasonable to be optimistic that, like all technology has been, it will be extraordinarily beneficial to humankind in the long term. To some degree, that takes care of itself. The short-term disruptions are why we should be informed about it, so we can know how to maximize the good and minimize the bad.

As Louis Pasteur—French chemist, microbiologist, and inventor of the principles behind vaccination and pasteurization—once said, "Chance favors only the prepared mind."[22] This book was neither meant to scare nor reassure. It's just meant to help you make up your own mind. I *do* personally hope you're more optimistic about the future now, but if nothing else, I hope you're more prepared.



With AI's rapid development, I wanted to provide a place for more resources, errata, and updates. For these and more, please visit the link in the QR code here.

ACKNOWLEDGMENTS

Writing your first book is like having your first child. You know it’s meant to be hard. When others do it, you make all of the appropriate noises about how big of an accomplishment it is. It must have been so difficult. How did you manage it? It’s not that these sentiments are insincere. It’s just impossible to fully appreciate the magnitude of the act without doing it. You can’t imagine the sheer number of people involved. Or the time truly required—that you already didn’t have.

As for how it compares to your second book or child—ask me later because I have neither right now.

An enormous number of individuals were instrumental to this book coming into being by giving early feedback, volunteering their expertise, and helping make the right connections to bring richness and depth to the text in these pages. An old saying goes, “Success has many fathers, but failure is an orphan.” These individuals all helped bring this book to life. For any inevitable errors and omissions, I claim paternity.

First off, thanks to the entire team at Manuscripts. They were excellent at shepherding me through what would have otherwise been a byzantine process. Their patience with a newcomer to the field and the professionalism they brought to this book cannot be overstated.

For providing the starting momentum on this project, I want to thank Brian Bies, whose dogged insistence I should write a book eventually overcame my objections that I didn’t have the time. This wouldn’t have happened without that encouragement. My gratitude as well to Eric Koester for his advice and starting this ball rolling down the hill for Manuscripts’ creation to begin with.

Thank you to my editors, Ken Cain and Ty Mall. Their many comments and revisions strained the limits of Google Docs in ways I could have never imagined. It even brought my Mac Studio to its knees—a machine that I've used to run some of the largest open-weight frontier LLMs. But that just speaks to the sheer magnitude of impact they had and the help I needed hammering the book into shippable form. Their involvement immeasurably improved the prose of this book, one comment or push to do better at a time.

Thanks also to George Thorne, whose extensive experience and understanding of the publishing industry significantly contributed to the success of the go-to-market strategy. A start-up entrepreneur and venture investor like me deeply appreciates the importance of this critical skill. And thanks to Shanna Heath, Mariana Montini, and Clayton Bohle for your guidance through the process. Finally, thank you to the broader support team at Manuscripts, whose efforts helped bring this book together.

When thinking about embarking on the journey of being an author, a special thanks to Evelyn Skye and Paul Podolsky. I very much appreciate their incredible generosity and willingness to carve out time to think through the process with me—in the midst of their own book launches! I'm deeply touched by their willingness to pause their hectic schedules and share their hard-earned insights with a first-time author. On a similar note, thank you to Stephanie Lim for helping provide her valuable perspective on the publishing industry. My neurotic self was able to take the plunge after getting my "really, this is the last" questions answered. All of them were so patient with me, no matter how ridiculous my scattered queries must have seemed.

I also want to give a special thanks to Chris Yeh. He also provided guideposts on the author's journey and gifted me a copy of the book he and Reid Hoffman wrote: *Blitzscaling*. This book was pivotal in surviving an unpleasantly eventful flight back to San Francisco and also demonstrated how one writes an engaging business book. His words of praise and thumbs-up to Reid were welcome encouragement to make my last push across the finish line.

Thank you as well to Sergei Polevikov and Devansh. I appreciate both of you constantly helping boost my book posts. Additionally, reading your work always makes me better informed and smarter. I'm glad to call you both friends after interacting with you so much on Substack and then finally meeting each of you in person.

Of course, this book could not have come together without the expertise of all those who generously volunteered their time to speak with me. Among those I can name, thank you: Alex Campbell, Alejandro Morffis, Ashok Goel, Brandon Basso, Devansh, Istvan Berkeley, Josh Blanchfield, Joshua Reicher, Michael Muelly, Rafic Makki, Tom Toner, and Tomide Adesanmi. Many more spoke anonymously, on background, or under Chatham House rules. You know who you are, and I thank you all so much. I have learned so much in the process of writing this book from everyone here.

To those who read my early drafts and provided testimonials, I'm deeply grateful. All of you have busy schedules, and to take the time to both read and provide thoughtful commentary is incredibly generous. I may have already expressed my gratitude to some of you in other sections, but you deserve it. Thank you, Alex Campbell, Ashok Goel, Brandon Reed, Devansh, Jordan Schneider, Josh Blanchfield, Joshua Reicher, Rafic Makki, Reid Hoffman, and Sergei Polevikov.

Cliché or not, my family played a critical role in this book as well. My wife Liz helped along every step of the way with marketing and wherever else she could. She's my partner in life but also in this book. Isabel provided examples and lived experience for what babies can or can't do, though she always leaned on the "can" side in surprising and (sometimes delightfully, sometimes not delightfully) unexpected ways. Thank you to my parents for providing encouragement every step of the way as well.

Finally, last but not least, thank you so much to my preorder supporters. As with many of these kinds of early-early preorder scenarios, there are many familiar names, including some that have appeared in this very acknowledgment already. However, I was surprised by the number of unfamiliar names that followed my work on Substack and other platforms. All of you took a leap of faith in me, which deeply touched me. Many of

you have said I’m too modest and shouldn’t be surprised, but I was still shocked by the number of books so many of you bought. Your confidence in a book that wasn’t yet produced but that you were already completely sure you wanted to share with friends and family helped drive me through the last, difficult stages in getting to the finish line. And yes, perhaps it’s *yet another* expression of too much modesty, but I thank you regardless of whether those copies actually make it into the hands of friends and family or if they collect dust in a box in the attic. All of you being so encouraging and casting those votes of confidence is what counts.

Lars Hallenius, Manuel Sescosse, Grant Hogarth, Josh Blanchfield, Travis Arnold, Chloe Alpert, Luca Schenk, Orn Orpong, Suresh Kumar Arjunan, Champ Suthipongchai, Peter Chau, Joy Wang, William Wang, Tami Flores, Betsy Valu Rohney, Miranda Tei, Nydia Zhang, Dina Nagao, Anna Lee, Razal Minhas, Behram Dacosta, Henry Heckman, Ryden Holman, Justin Yi, John McKenna, Seferina Klinger, Theo Saarinen, James Hands, Douglas Pardo, Evelyn Skye, Justin Ash, Jessica Karr, Jerry Hamm, Joe Kersjes, Martin Travis, Eric Koester, Ee Fann Goh, Jason Stinson, Sam Spink, Sarah Chen, Harmen Van der Linde, and Robbie Hume.

Regardless of how many people ultimately buy the book, you were all the earliest and helped make this book happen. For that, I am eternally grateful.

NOTES

INTRODUCTION

1. Dan Milmo, "'Godfather of AI' Shortens Odds of the Technology Wiping out Humanity over next 30 Years," *Artificial Intelligence* (blog), *The Guardian*, December 24, 2024, https://www.theguardian.com/technology/2024/dec/27/godfather-of-ai-raises-odds-of-the-technology-wiping-out-humanity-over-next-30-years.
2. Marc Andreessen, "Why AI Will Save the Word," *a16z* (blog), *Andreessen Horowitz*, June 3, 2023, https://a16z.com/ai-will-save-the-world/.
3. Yun Li, "Brad Gerstner Says AI Will Be Bigger than the Internet, Bigger than Mobile," *Delivering Alpha* (blog), *CNBC*, September 28, 2023, https://www.cnbc.com/2023/09/28/brad-gerstner-says-ai-is-going-to-be-bigger-than-the-internet-bigger-than-mobile.html.
4. Allison Nathan, Jenny Grimberg, and Ashley Rhodes, "Gen AI: Too Much Spend, Too Little Benefit?" *Goldman Sachs* no. 129 (June 2024), https://www.goldmansachs.com/images/migrated/insights/pages/gs-research/gen-ai--too-much-spend,-too-little-benefit-/TOM_AI%202.0_ForRedaction.pdf.
5. OpenAI, "Introducing ChatGPT," last modified November 30, 2022, https://openai.com/index/chatgpt/.
6. Jamie Ballard, "Americans' Top Feeling about AI: Caution," *Technology* (blog), *YouGov,* April 9, 2024, https://today.yougov.com/technology/articles/49099-americans-2024-poll-ai-top-feeling-caution.

CHAPTER 1

1. David Alan Grier, “Human Computers: The First Pioneers of the Information Age,” *Endeavour* 25, no. 1 (March 2001): 28–32, https://doi.org/10.1016/S0160-9327(00)01338-7.
2. Ford Motor Company, “The Model T,” *History* (blog), *Ford Motor Company*, 2024, Accessed March 17, 2025, https://corporate.ford.com/articles/history/the-model-t.html.
3. Associated Press, “Ford Motor Loses $3.1 Billion Due to Chip Shortage and Rivian,” *Moneywatch* (blog), *CBS News*, April 27, 2022, https://www.cbsnews.com/news/ford-motor-losses-chip-shortage-rivian/.
4. Steven Sass, “The US Professional Sector: 1950 to 1988,” N*ew England Economic Review* (January 1990): 37–55, https://ideas.repec.org/a/fip/fedbne/y1990ijanp37-55.html.
5. National Science Board, “Science and Engineering Indicators: The STEM Labor Force: Scientists, Engineers, and Skilled Technical Workers,” *Science & Engineering Indicators* (blog), *National Science Foundation*, May 30, 2024, https://ncses.nsf.gov/pubs/nsb20245.
6. The Editors of Encyclopædia Britannica, (Chicago, IL: Encyclopedia Britannica, 2025), s.v. “Industrial Revolution,” https://www.britannica.com/event/Industrial-Revolution.
7. John Rafferty, “The Rise of the Machines: Pros and Cons of the Industrial Revolution,” *History & Society* (blog), *Encyclopedia Britannica*, September 30, 2017, https://www.britannica.com/story/the-rise-of-the-machines-pros-and-cons-of-the-industrial-revolution.
8. Jason Crawford, “Why Did We Wait So Long for the Threshing Machine?” *Roots of Progress* (blog), June 29, 2021, https://blog.rootsofprogress.org/why-did-we-wait-so-long-for-the-threshing-machine.
9. John Rafferty, “The Rise of the Machines: Pros and Cons of the Industrial Revolution,” *History & Society* (blog), *Encyclopedia Britannica*, September 30, 2017, https://www.britannica.com/story/the-rise-of-the-machines-pros-and-cons-of-the-industrial-revolution.
10.

a. Max Roser, Bertha Rohenkohl, Pablo Arriagada, Joe Hasell, Hannah Ritchie, and Esteban Ortiz-Ospina, “Global GDP over the Long Run,” *Data* (blog), *Our World in Data*, 2023, Accessed May 12, 2025, https://ourworldindata.org/grapher/global-gdp-over-the-long-run.
b. Ben Wattenberg, “Infant Mortality and Life Expectancy,” *Timeline* (blog), *PBS*, n.d., accessed March 17, 2025, https://www.pbs.org/fmc/timeline/dmortality.htm.
c. The World Health Organization (WHO), “Child Mortality and Causes of Death,” *Topics* (blog), *WHO*, 2025, accessed March 17, 2025, https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death.
d. Bastian Herre, “Global Average Life Expectancy Has More than Doubled since 1900,” *Data Insights* (blog), *Our World in Data*, February 3, 2025, https://ourworldindata.org/data-insights/global-average-life-expectancy-has-more-than-doubled-since-1900.

11. Dictionary.com (Oakland, CA: Dictionary.com, 2025), s.v. “Luddite,” https://www.dictionary.com/browse/luddite.

12. Evan Andrews, “Who Were the Luddites?” *The History Channel* (blog), *History.com*, August 21, 2023, https://www.history.com/news/who-were-the-luddites.

13. Joseph Park, “The New Luddites,” *Blog* (blog), *Digital Architecture Lab*, May 4, 2024, https://dalab.xyz/en/blog/the-new-luddites/.

14. Evan Andrews, “Who Were the Luddites?” *Industrial Revolution* (blog), *The History Channel*, August 21, 2023, https://www.history.com/news/who-were-the-luddites.

15. Joseph Park, “The New Luddites,” *Blog* (blog), *Digital Architecture Lab*, May 4, 2024, https://dalab.xyz/en/blog/the-new-luddites/.

16.

a. Fred Shannon, “The Status of the Midwestern Farmer in 1900,” *The Mississippi Valley Historical Review* 37, no. 3 (December 1950): 491, https://doi.org/10.2307/1893323.

b. United States Department of Agriculture—National Agricultural Statistics Service, “Southern Region News Release Farm Labor” (Athens, GA: USDA, 2023), https://www.nass.usda.gov/Statistics_by_State/Regional_Office/Southern/includes/Publications/Economic_and_Demographic_Releases/Farm_Labor/2023/OctoberLabor2023.pdf.
c. Rural Migration News, “Agriculture: A Century of Change,” *Rural Migration Blog* (blog), *The Regents of the University of California*, April 25, 2018, https://migration.ucdavis.edu/rmn/blog/post/?id=2157.

17. United States Department of Agriculture (USDA), “Farming and Farm Income,” *Ag and Food Statistics: Charting the Essentials* (blog), *USDA*, March 12, 2025, https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/farming-and-farm-income.

18. SoA Policy Team, “SoA Survey Reveals a Third of Translators and Quarter of Illustrators Losing Work to AI,” *News* (blog), *The Society of Authors*, April 11, 2024, https://societyofauthors.org/2024/04/11/soa-survey-reveals-a-third-of-translators-and-quarter-of-illustrators-losing-work-to-ai/.

19. Ole Teutloff, Johanna Einsiedler, Otto Kässi, et al. “Winners and Losers of Generative AI: Early Evidence of Shifts in Freelancer Demand,” *Journal of Economic Behavior & Organization* 235 (July 2025): 106845, https://doi.org/10.1016/j.jebo.2024.106845.

20. AIPRM, “AI in the Workplace Statistics 2024,” *Featured* (blog), *AIPRM*, 2025, accessed May 12, 2025, https://www.aiprm.com/ai-in-workplace-statistics/.

21. Pew Research, “US Workers Are More Worried Than Hopeful about Future AI Use in the Workplace,” *Report* (blog), *Pew Research Center*, February 25, 2025, https://www.pewresearch.org/social-trends/2025/02/25/u-s-workers-are-more-worried-than-hopeful-about-future-ai-use-in-the-workplace/.

22. Molly Kinder, Xavier de Souza Briggs, Mark Muro, and Sifan Liu, “Generative AI, the American Worker, and the Future of Work,”

Research (blog), *Brookings Institute*, October 10, 2024, https://www.brookings.edu/articles/generative-ai-the-american-worker-and-the-future-of-work/.

23. Donald MacRae, (Chicago, IL: Encyclopedia Britannica, 2025), s.v. "Thomas Malthus," https://www.britannica.com/money/Thomas-Malthus.

24. United Nations, "Population," *Global Issues* (blog), n.d., accessed March 19, 2025, https://www.un.org/en/global-issues/population.

25. Jeremy Rifkin, *The End of Work: The Decline of the Global Labor Force and the Dawn of the Post-Market Era* (New York: Putnam, 1996): 288–290.

26. Federal Reserve Bank of St. Louis, "FRED Graph," n.d., accessed March 19, 2025. https://fred.stlouisfed.org/graph/?g=1EXqh.

27. Scott Lincicome and Arjun Anand, "The 'China Shock' Demystified: Its Origins, Effects, and Lessons for Today," *Defending Globalization: Economics* (blog), *Cato Institute*, December 12, 2023, https://www.cato.org/publications/china-shock.

28. Femke Nijsse, Jean-Francois Mercure, Nadia Ameli, Francesca Larosa, Sumit Kothari, Jamie Rickman, Pim Vercoulen, and Hector Pollitt, "The Momentum of the Solar Energy Transition," *Nature Communications* 14, no. 1 (October 2023): 6542, https://doi.org/10.1038/s41467-023-41971-7.

29. Indeed.com, "Autonomous Driving Engineer Jobs in California," 2025, accessed September 5, 2024, https://www.indeed.com/q-autonomous-driving-engineer-l-california-jobs.html?vjk=83727683ce8726da.

CHAPTER 2

1. Harlan Ellison, *Greatest Hits (Herald Classics)* (New York: Union Square & Co., 2024), 19–37.

2. Ethan Shanfeld, "Joe Biden Grew More Worried about AI after Watching 'Mission: Impossible—Dead Reckoning,' Says White House

Deputy," *Film News* (blog), *Variety*, October 31, 2023, https://variety.com/2023/film/news/joe-biden-worried-ai-mission-impossible-dead-reckoning-1235775115/.

CHAPTER 3

1. Jo Ann Oravec, "Expert Systems and Knowledge-Based Engineering (1984–1991): Implications for Instructional Systems Research," *International Journal of Designs for Learning* 5, no. 2 (November 2014), https://doi.org/10.14434/ijdl.v5i2.12891.
2. Edward Feigenbaum, Pamela McCorduck, and H. Penny Nii, *The Rise of the Expert Company: How Visionary Companies Are Using Artificial Intelligence to Achieve Higher Productivity and Profits* (London: Macmillan London, 1988), 9.
3. Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Hoboken, NJ: Pearson, 2021), 24.
4. T. Menzies, "21st-Century AI: Proud, Not Smug," *IEEE Intelligent Systems* 18, no. 3 (May 2003): 19, https://doi.org/10.1109/MIS.2003.1200723.
5. Ibid.
6. Edward Feigenbaum, Pamela McCorduck, and H. Penny Nii, *The Rise of the Expert Company: How Visionary Companies Are Using Artificial Intelligence to Achieve Higher Productivity and Profits* (London: Macmillan London, 1988), 9–12.
7. Sebastian Schuchmann, "Analyzing the Prospect of an Approaching AI Winter," (bachelor's thesis, Monash University, Australia, 2019), https://doi.org/10.13140/RG.2.2.10932.91524.
8. Elena Benderskaya and Sofya Zhukova, "Multidisciplinary Trends in Modern Artificial Intelligence: Turing's Way," in *Artificial Intelligence, Evolutionary Computing and Metaheuristics* 427, ed. Xin-She Yang (Berlin, Heidelberg: Springer Berlin Heidelberg, 2013): 319–343, https://doi.org/10.1007/978-3-642-29694-9_13.

9. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Hoboken, NJ: Pearson, 2021), 222–237.

10. Reid Smith, “Knowledge-Based Systems: Concepts, Techniques, Examples,” (Lansdowne Park, Ottawa, ON, 1985).

11. John Haugeland, *Artificial Intelligence: The Very Idea* (Cambridge, MA: MIT Press, 1989), 112, https://doi.org/10.7551/mitpress/1170.001.0001.

12. Elena Benderskaya and Sofya Zhukova, “Multidisciplinary Trends in Modern Artificial Intelligence: Turing’s Way,” in *Artificial Intelligence, Evolutionary Computing and Metaheuristics* 427, ed. Xin-She Yang (Berlin, Heidelberg: Springer Berlin Heidelberg, 2013): 319–343, https://doi.org/10.1007/978-3-642-29694-9_13.

13. David Stork, “HAL’s Legacy: Two; Scientist on the Set: An Interview with Marvin Minsky,” *MIT Press Scientist on the Set*, last modified January 14, 1998, https://web.archive.org/web/20120716182537/http://mitpress.mit.edu/e-books/Hal/chap2/two2.html.

14. David Rumelhart, James McClelland, and PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations* 1 (Cambridge, MA: MIT Press, 1986), 43, https://doi.org/10.7551/mitpress/5236.001.0001.

15. Mikel Olazaran, “A Sociological Study of the Official History of the Perceptrons Controversy,” *Social Studies of Science* 26, no. 3 (August 1996): 611–659, https://doi.org/10.1177/030631296026003005.

16. Sebastian Schuchmann, “Analyzing the Prospect of an Approaching AI Winter,” (bachelor’s thesis, Darmstadt University of Applied Sciences, 2019), https://doi.org/10.13140/RG.2.2.10932.91524.

17. Stephen Wolfram, “What Is ChatGPT Doing… and Why Does It Work?” *Writings* (blog), *Stephen Wolfram*, February 14, 2023, https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

18. Alex Kiefer, “A Defense of Pure Connectionism,” (PhD thesis, Monash University, Australia, 2019),

https://doi.org/10.13140/RG.2.2.18476.51842.

19.

a. Warren McCulloch and Walter Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *The Bulletin of Mathematical Biophysics* 5, no. 4 (December 1943): 115–133, https://doi.org/10.1007/BF02478259.

b. Akshay Chandra, “McCulloch-Pitts Neuron—Mankind’s First Mathematical Model of a Biological Neuron,” *TDS Archive* (blog), *Medium*, July 24, 2018, https://medium.com/data-science/mcculloch-pitts-model-5fdf65ac5dd1.

20. Frank Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” *Psychological Review* 65, no. 6 (1958): 386–408, https://doi.org/10.1037/h0042519.

21. Melanie Lefkowitz, “Professor’s Perceptron Paved the Way for AI—60 Years Too Soon,” *Stories* (blog), *Cornell News*, September 25, 2019, https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon.

22. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning,” *Nature* 521, no. 7553 (May 2015): 436–444, https://doi.org/10.1038/nature14539.

23. Melanie Lefkowitz, “Professor’s Perceptron Paved the Way for AI—60 Years Too Soon,” *Stories* (blog), *Cornell News*, September 25, 2019, https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon.

24. Ibid.

25. Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, An A K Peters Book (Boca Raton, FL: CRC Press, 2018) 87.

26. Stephen Wolfram, “Farewell, Marvin Minsky (1927-2016)” *Writings* (blog), *Stephen Wolfram*, January 26, 2016, https://writings.stephenwolfram.com/2016/01/farewell-marvin-minsky-19272016/.

27. Melanie Lefkowitz, “Professor’s Perceptron Paved the Way for AI—60 Years Too Soon,” *Stories* (blog), *Cornell News*, September 25, 2019, https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon.

28. Claude Cruz, *Understanding Neural Networks: A Primer* (Amherst, NH: Graeme, 1988), 2.

29. Yann Gourvennec, “Perceptron: A History of Neural Networks,” *AI and Big Data* (blog), *Visionary Marketing & Innovation*, April 22, 2024, https://visionarymarketing.com/en/2024/04/22/perceptron-a-history-of-neural-networks/.

30. Frank Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” *Psychological Review* 65, no. 6 (1958): 386–408, https://doi.org/10.1037/h0042519.

31. Warren S. McCulloch and Walter Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *The Bulletin of Mathematical Biophysics* 5, no. 4 (December 1943): 115–133, https://doi.org/10.1007/BF02478259.

32. Melanie Lefkowitz, “Professor’s Perceptron Paved the Way for AI—60 Years Too Soon,” *Stories* (blog), *Cornell News*, September 25, 2019, https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon.

33. Frank Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” *Psychological Review* 65, no. 6 (1958): 386–408, https://doi.org/10.1037/h0042519.

34. Melanie Lefkowitz, “Professor’s Perceptron Paved the Way for AI—60 Years Too Soon,” *Stories* (blog), *Cornell News*, September 25, 2019, https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon.

35. Mikel Olazaran, “A Sociological Study of the Official History of the Perceptrons Controversy,” *Social Studies of Science* 26, no. 3 (August 1996): 342–343, https://doi.org/10.1177/030631296026003005.

36. Cornell Daily Sun Staff, “Prof. Rosenblatt Dies,” *The Cornell Daily Sun*, 88th ed., September 1, 1971.

37. Jon Agar, “What Is Science For? The Lighthill Report on Artificial Intelligence Reinterpreted,” *The British Journal for the History of Science* 53, no. 3 (September 2020): 289–310, https://doi.org/10.1017/S0007087420000230.

38. Sebastian Schuchmann, “Analyzing the Prospect of an Approaching AI Winter,” (bachelor’s thesis, Monash University, Australia, 2019), https://doi.org/10.13140/RG.2.2.10932.91524.

39. Marvin Minsky and Seymour Papert, *Perceptrons: An Introduction to Computational Geometry*, Expanded ed. (Cambridge, MA: MIT Press, 1988).

40. Sean Trott, “Perceptrons, XOR, and the First ‘AI Winter,’” *The Counterfactual* (blog), *Sean Trott*, January 17, 2024, https://seantrott.substack.com/p/perceptrons-xor-and-the-first-ai.

41.

    a. Claude Cruz, *Understanding Neural Networks: A Primer* (Amherst, NH: Graeme, 1988).

    b. NobelPrize.org, “John J. Hopfield—Facts—2024,” *Physics* (blog), *Nobel Prize Outreach 2025*, March 31, 2025, https://www.nobelprize.org/prizes/physics/2024/hopfield/facts/.

42. Edward Feigenbaum, *The Art of Artificial Intelligence: I. Themes and Case Studies of Knowledge Engineering* (Stanford, CA: Stanford University, 1977).

43. Cornelius Leondes, *Expert Systems: The Technology of Knowledge Management and Decision Making for the 21st Century* (San Diego, CA: Academic Press, 2002).

44. Mark Stefik, *Introduction to Knowledge Systems* (San Francisco, CA: Morgan Kaufmann Publishers, 1995), 365–366.

45. Ibid, 370–375.

46. Huzaifa Sidpurwala, “When LLMs Day Dream: Hallucinations and How to Prevent Them,” *The Red Hat Blog* (blog), *Red Hat*, September 25, 2024, https://www.redhat.com/en/blog/when-llms-day-dream-hallucinations-how-prevent-them.

47. Jay Liebowitz, “Common Fallacies about Expert Systems,” *ACM SIGCAS Computers and Society* 16–17, no. 4–1 (October 1987): 28–33, https://doi.org/10.1145/25601.25605.

48. Steven Borowiec, “AlphaGo Seals 4–1 Victory over Go Grandmaster Lee Sedol,” *Technology* (blog), *The Guardian*, March 15, 2016, https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol.

49. Istvan Berkeley, “The Curious Case of Connectionism,” *Open Philosophy* 2, no. 1 (January 2019): 190–205, https://doi.org/10.1515/opphil-2019-0018.

CHAPTER 4

1. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Hoboken, NJ: Pearson, 2021), 676–678.

2. Jon Agar, “What Is Science for? The Lighthill Report on Artificial Intelligence Reinterpreted,” *The British Journal for the History of Science* 53, no. 3 (September 2020): 289–310, https://doi.org/10.1017/S0007087420000230.

3. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning,” *Nature* 521, no. 7553 (May 2015): 436, https://doi.org/10.1038/nature14539.

4. Ibid.

5. Sean Trott, “Perceptrons, XOR, and the First ‘AI Winter,’” *The Counterfactual* (blog), *Substack*, January 17, 2024, https://seantrott.substack.com/p/perceptrons-xor-and-the-first-ai.

6.

a. Kurt Hornik, Maxwell Stinchcombe, and Halbert White, “Multilayer Feedforward Networks Are Universal Approximators,” *Neural Networks* 2, no. 5 (1989): 1, https://doi.org/10.1016/0893-6080(89)90020-8.
b. Geeks for Geeks, “Universal Approximation Theorem for Neural Networks,” last modified September 24, 2024, https://www.geeksforgeeks.org/universal-approximation-theorem-for-neural-networks/.

7. David Rumelhart, James McClelland, and PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations* (Cambridge, MA: MIT Press, 1986), 533, https://doi.org/10.7551/mitpress/5236.001.0001.
8. Ibid, 534–535.
9. Paul Werbos, “Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences,” (PhD dissertation, Harvard University, 1974), https://www.researchgate.net/publication/35657389_Beyond_regression_new_tools_for_prediction_and_analysis_in_the_behavioral_sciences.
10. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning,” *Nature* 521, no. 7553 (May 2015): 438, https://doi.org/10.1038/nature14539.
11. Jürgen Schmidhuber, “Deep Learning in Neural Networks: An Overview,” *Neural Networks* 61 (2015): 94, https://doi.org/10.1016/j.neunet.2014.09.003.
12. Olga Russakovsky, Jia Deng, Hao Su, et al., “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision* 115, no. 3 (2015): 211–252, https://doi.org/10.1007/s11263-015-0816-y.
13. Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM* 60, no. 6 (2017): 84–90,

https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

14. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep Learning,” *Nature* 521, no. 7553 (May 2015): 440, https://doi.org/10.1038/nature14539.

15. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016): 770–778, https://doi.org/10.1109/CVPR.2016.90.

16. Peter Eckersley and Yomna Nasser, “EFF AI Progress Measurement Project,” Electronic Frontier Foundation, last modified 2017, accessed September 7, 2018, https://web.archive.org/web/20170929190726/https://eff.org/ai/metrics.

17. Netflix Technology Blog, “Netflix Recommendations: Beyond the 5 stars (Part 2),” *Netflix TechBlog* (blog), *Medium*, June 20, 2012, https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5.

18. George Stratton, “Vision without Inversion of the Retinal Image,” *Psychological Review* 3, no. 6 (1896): 611–617, https://psycnet.apa.org/record/1926-02881-001.

19. J. Lettvin, H. Maturana, W. McCulloch, and W. Pitts, “What the Frog’s Eye Tells the Frog’s Brain,” *Proceedings of the IRE* 54, no. 11 (1966): 1940–1951, https://hearingbrain.org/docs/letvin_ieee_1959.pdf.

20. Frank Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” *Psychological Review* 65, no. 6 (1958): 388, https://doi.org/10.1037/h0042519.

21. Peter Eckersley and Yomna Nasser, “EFF AI Progress Measurement,” Electronic Frontier Foundation, last modified 2017, accessed September 7, 2018, https://web.archive.org/web/20170929190726/https://eff.org/ai/metrics.

22. NobelPrize.org, “John J. Hopfield,” *The Nobel Prize in Physics* (blog), *The Nobel Prize*, March 24, 2025,

https://www.nobelprize.org/prizes/physics/2024/hopfield/facts/.

23. Dessa, “DeepLearni.ng Is History,” *Dessa* (blog), *Medium*, June 11, 2018, https://medium.com/@dessa_/deeplearni-ng-is-dessa-2cd8f4b0fc4d.

24. Anna Tong and Akanksha Khushi, “OpenAI Co-Founder Ilya Sutskever Departs ChatGPT Maker,” *Technology* (blog), *Reuters*, May 14, 2024, https://www.reuters.com/technology/openai-co-founder-ilya-sutskever-departs-2024-05-14/.

25. Stephen Wolfram, “What Is ChatGPT Doing… and Why Does It Work?” *Writings* (blog), *Stephen Wolfram*, February 14, 2023, https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.

26.

    a. Sundar Pichai and Demis Hassabis, “Our Next-Generation Model: Gemini 1.5,” *Technology* (blog), *Google*, February 15, 2024, https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/.

    b. Hugo Touvron, Thibaut Lavril, Gautier Izacard, et al., “LLaMA: Open and Efficient Foundation Language Models,” *arXiv* (February 2023), https://doi.org/10.48550/ARXIV.2302.13971.

    c. Jack Lindsey, Wes Gurnee, Emmanuel Ameisen, et al., “On the Biology of a Large Language Model,” *Transformer Circuits Thread* (March 2025), https://transformer-circuits.pub/2025/attribution-graphs/biology.html.

    d. DeepSeek-AI, Aixin Liu, Bei Feng, et al., “DeepSeek-V3 Technical Report,” *arXiv* (December 2024), https://doi.org/10.48550/ARXIV.2412.19437.

27. Jamie Sandhu, “What Are LLMs and Generative AI? A Beginner’s Guide to the Technology Turning Heads,” *Commentary* (blog), *University of Toronto*, January 25, 2024, https://srinstitute.utoronto.ca/news/gen-ai-llms-explainer.

28. Ibid.

29. Andrew Ng and Michael Jordan, “On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes,” in *Proceedings of the 15th International Conference on Neural Information Processing Systems: Natural and Synthetic* (Cambridge, MA: MIT Press, 2001), 841–848.

30. Thorsten Brants, Ashok Popat, Peng Xu, et al., “Large Language Models in Machine Translation,” in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (2007): 858–867.

31. Ashish Vaswani, Noam Shazeer, Niki Parmar, et al., “Attention Is All You Need,” *arXiv* (June 2017), https://doi.org/10.48550/ARXIV.1706.03762.

32. Ibid, 6.

33. Ashish Vaswani, Noam Shazeer, Niki Parmar, et al., “Attention Is All You Need,” *arXiv* (June 2017), 6, https://doi.org/10.48550/ARXIV.1706.03762.

34. Alex Reisner, “The Unbelievable Scale of AI’s Pirated-Books Problem,” *Technology* (blog), *The Atlantic*, March 20, 2025, https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/.

CHAPTER 5

1. Fabio Duarte, “Number of ChatGPT Users,” *Exploding Topics* (blog), March 25, 2025, https://explodingtopics.com/blog/chatgpt-users.

2. @alberta.tech, “Why Can No AI Answer: How Many Rs Are in Strawberry?” *Instagram Reel*, August 26, 2024, https://www.instagram.com/reel/C_IkB-cuCZF/.

3. David Cole, *The Stanford Encyclopedia of Philosophy*, winter 2024 Ed. (Standford, CA: The Metaphysics Research Lab, Department of Philosophy, Stanford University, 2024), s.v. “The Chinese Room

Argument," https://plato.stanford.edu/archives/win2024/entries/chinese-room/.

4. Jason Chein, Steven Martinez, and Alexander Barone, "Human Intelligence Can Safeguard Against Artificial Intelligence: Individual Differences in the Discernment of Human from AI Texts," *Scientific Reports* 14 (October 2024): 25989, https://doi.org/10.1038/s41598-024-76218-y.
5. Jenna Russell, Marzena Karpinska, and Mohit Iyyer, "People Who Frequently Use ChatGPT for Writing Tasks Are Accurate and Robust Detectors of AI-Generated Text," *arXiv* (January 2025), https://doi.org/10.48550/ARXIV.2501.15654.
6. Google, "What Is Artificial General Intelligence (AGI)," *AGI* (blog), *Google Cloud*, n.d., accessed April 14, 2025, https://cloud.google.com/discover/what-is-artificial-general-intelligence.
7. Gary Marcus, "AI Has (Sort of) Passed the Turing Test; Here's Why That Hardly Matters," *Marcus on AI* (blog), *Substack*, April 2, 2025, https://garymarcus.substack.com/p/ai-has-sort-of-passed-the-turing.
8. News, "AI Is Dumber Than a Cat? That's What an AI Pioneer Thinks," *Briefs* (blog), *Mind Matters*, October 14, 2024, https://mindmatters.ai/brief/ai-is-dumber-than-a-cat-thats-what-an-ai-pioneer-thinks/.
9. Benj Edwards, "OpenAI Experiments with Giving ChatGPT a Long-Term Conversation Memory," *Informed Technology* (blog), *Ars Technica*, February 13, 2024, https://arstechnica.com/information-technology/2024/02/amnesia-begone-soon-chatgpt-will-remember-what-you-tell-it-between-sessions/.
10. Samuel Axon, "ChatGPT Can Now Remember and Reference All Your Previous Chats," *LLMs* (blog), *Ars Technica*, April 10, 2025, https://arstechnica.com/ai/2025/04/chatgpt-can-now-remember-and-reference-all-your-previous-chats/.
11. Ashok Goel, "Looking Back, Looking Ahead: Symbolic versus Connectionist AI," *AI Magazine* 42, no. 4 (2021): 83–85,

https://doi.org/10.1609/aaai.12026.

12. Ludwig Wittgenstein, “Chomsky and Universal Grammar,” *Bad Language* (blog), *The University of British Columbia*, December 8, 2022, https://blogs.ubc.ca/badlanguage/2019/06/20/chomsky-and-universal-grammar/.

13. Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Proceedings of the 26th International Conference on Neural Information Processing Systems—Volume 1*, NIPS’12 (Red Hook, NY: Curran Associates Inc., 2012), 1,097–1,105, https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

14. François Chollet, “On the Measure of Intelligence,” *arXiv* (November 2019), 13, https://doi.org/10.48550/ARXIV.1911.01547.

15. François Chollet, “OpenAI o3 Breakthrough High Score on ARC-AGI-PUB,” *ARC Prize* (blog), December 20, 2024, https://arcprize.org/blog/oai-o3-pub-breakthrough.

16. Jason Wei, Xuezhi Wang, Dale Schuurmans, et al., “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models,” *arXiv* (January 2023), https://doi.org/10.48550/ARXIV.2201.11903.

17. OpenAI, “Learning to Reason with LLMs,” *Release* (blog), *OpenAI*, September 12, 2024, https://openai.com/index/learning-to-reason-with-llms/.

18. James Wang, “What Mattered in GenAI in 2024,” *Weighty Thoughts* (blog), *Substack*, March 12, 2025, https://weightythoughts.com/p/what-mattered-in-genai-in-2024.

19. Kyle Wiggers, “OpenAI’s o3 Model Might Be Costlier to Run than Originally Estimated,” *TechCrunch*, April 2, 2025, https://techcrunch.com/2025/04/02/openais-o3-model-might-be-costlier-to-run-than-originally-estimated/.

20. James Wang, “AI Reasoning—What Is It?” *Weighty Thoughts* (blog), *Substack*, January 17, 2025, https://weightythoughts.com/p/ai-

reasoningwhat-is-it.

21. Marc Andreessen and Ben Horowitz, “The State of AI with Marc & Ben,” *a16z Podcast*, released June 14, 2024, 01:19:49, https://a16z.com/podcast/the-state-of-ai-with-marc-ben/.

22. Logan Thorneloe, “ML for SWES Weekly #11: The Engineer’s Perspective on Apple’s LLM Reasoning Paper,” *Machine Learning for Software Engineers* (blog), *Substack*, https://mlforswes.com/p/6-10-25.

23. James Wang, “AI Reasoning—What Is It?” *Weighty Thoughts* (blog), *Substack*, January 17, 2025, https://weightythoughts.com/p/ai-reasoningwhat-is-it.

24. Greg Kamradt, “ARC-AGI-2 + ARC Prize 2025 Is Live!” *ARC Prize* (blog), March 24, 2025, https://arcprize.org/blog/announcing-arc-agi-2-and-arc-prize-2025.

25. Grace Zhao, “Stanford Misinformation Expert Admits to ChatGPT ‘Hallucinations’ in Court Statement,” *Science & Technology* (blog), *The Stanford Daily*, December 4, 2024, https://stanforddaily.com/2024/12/04/hancock-admitted-to-ai-use/.

26. Christopher Kohls and Mary Franson v. Keith Ellison and Chad Larson, 24-cv-03754 (LMP/DLM), “Expert Declaration of Professor Jeff Hancock” (United States District Court, District of Minnesota. 2024), https://storage.courtlistener.com/recap/gov.uscourts.mnd.220348/gov.uscourts.mnd.220348.23.0.pdf.

27. Sara Merken, “AI ‘Hallucinations’ in Court Papers Spell Trouble for Lawyers,” *Artificial Intelligence* (blog), *Reuters*, February 18, 2025, https://www.reuters.com/technology/artificial-intelligence/ai-hallucinations-court-papers-spell-trouble-lawyers-2025-02-18/.

28. Huzaifa Sidpurwala, “When LLMs Day Dream: Hallucinations and How to Prevent Them,” *Red Hat Blog* (blog), *Red Hat*, September 25, 2024, https://www.redhat.com/en/blog/when-llms-day-dream-hallucinations-how-prevent-them.

29. Emily Bender, Timnit Gebru, Angelina McMillan-Major, et al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” in

Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM, 2021): 610–23, https://doi.org/10.1145/3442188.3445922.

CHAPTER 6

1. Eric Wallace, Florian Tramèr, Matthew Jagielski, and Ariel Herbert-Voss, "Does GPT-2 Know Your Phone Number?" *Berkeley Artificial Intelligence Research* (blog), *University of California Berkeley*, December 20, 2020, http://bair.berkeley.edu/blog/2020/12/20/lmmem/.
2. Rachel Reed, "ChatNYT," *National & World Affairs* (blog), *Harvard Law School*, March 22, 2024, https://hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/.
3. Jason Coombs, "Understanding Oracle Attacks on Information Services," *Dr. Dobb's* (blog), *Informa Tech*, December 2, 2004, http://www.drdobbs.com/understanding-oracle-attacks-on-informat/184405917.
4. Anne Helmenstine, "How Many Atoms Exist in the Universe?" *Chemistry* (blog), *ThoughtCo.*, August 8, 2019, https://www.thoughtco.com/number-of-atoms-in-the-universe-603795.
5. OpenAI, "Introducing Deep Research," *Release* (blog), February 2, 2025, https://openai.com/index/introducing-deep-research/.
6. Ben Thompson, "Deep Research and Knowledge Value," *Stratechery* (blog), February 10, 2025, https://stratechery.com/2025/deep-research-and-knowledge-value/.
7. James Wang and Devansh, "How Will AI Impact Law Firms," *Weighty Thoughts* (blog), *Substack*, February 8, 2025, https://weightythoughts.com/p/how-will-ai-impact-law-firms.
8. B. Sullivan, "'Any Color the Customer Wants, as Long as It's Black.' Henry Ford," James Madison Museum of Orange County Heritage, last modified February 15, 2023,

https://www.thejamesmadisonmuseum.net/single-post/any-color-the-customer-wants-as-long-as-it-s-black-henry-ford.

9. Eman Elshaikh, “What is the Green Revolution? | World History Project,” OER Project, November 13, 2019, 00:09:59, https://www.youtube.com/watch?v=015HH9zbWHQ.

10. Jeff Smoot, “The History of USB Standards from 1.0 to USB4,” *Same Sky* (blog), June 2, 2022, https://www.sameskydevices.com/blog/the-history-of-usb-standards-from-1-to-usb4.

11. Elizabeth Van Arsdall, “Never Doubt the Power of Patterns,” *CogBlog* (blog), *Colby College*, November 30, 2020, https://web.colby.edu/cogblog/2020/11/30/never-doubt-the-power-of-patterns/.

12. Mehgan Andrade and Neil Walker, “12.1: Approaches to Pattern Recognition,” *Cognitive Psychology* (blog), *LibreTexts Social Sciences*, January 2, 2025, https://socialsci.libretexts.org/Bookshelves/Psychology/Cognitive_Psychology/Cognitive_Psychology_%28Andrade_and_Walker%29/12%3A_Classification_and_Categorization_with_Pattern_Recognition/12.01%3A_Approaches_to_Pattern_Recognition.

13. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (December 2016): 770–778, https://doi.org/10.1109/CVPR.2016.90.

14. Yonghui Wu, Mike Schuster, Zhifeng Chen, et al., “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation,” *arXiv* (October 2016), https://doi.org/10.48550/ARXIV.1609.08144.

15. Everton Aleixo, Juan Colonna, Marco Cristo, et al., “Catastrophic Forgetting in Deep Learning: A Comprehensive Taxonomy,” *arXiv*, (December 2023), https://doi.org/10.48550/arXiv.2312.10549.

16. Naftali Tishby and Noga Zaslavsky, “Deep Learning and the Information Bottleneck Principle,” *arXiv* (March 2015),

https://doi.org/10.48550/ARXIV.1503.02406.

17. Ernest Davis and Gary Marcus, “Reports of LLMs mastering math have been greatly exaggerated,” *Marcus on AI* (blog), *Substack*, April 5, 2025, https://garymarcus.substack.com/p/reports-of-llms-mastering-math-have.

18. Ibid.

19. Amazon, “What Is Reinforcement Learning?” *Machine Learning* (blog), *Amazon Web Services*, 2025, https://aws.amazon.com/what-is/reinforcement-learning/.

20. Peter Whidden, “Training AI to Play Pokemon with Reinforcement Learning,” peterwhidden, October 8, 2023, 33:52:00, https://www.youtube.com/watch?v=DcYLT37ImBY.

21. Nathan Lambert, “A Recipe for Frontier Model Post-Training,” *Interconnects* (blog), *Substack*, August 7, 2024, https://www.interconnects.ai/p/frontier-model-post-training.

22. @ruoyuGao, @PaparaZz1, and @hijkzzz, et al., “README.md,” *awesome-RLHF* (blog), *Github*, February 18, 2025, https://github.com/opendilab/awesome-RLHF/blob/main/README.md.

23. IBM, “Supervised versus Unsupervised Learning: What’s the Difference?” *Artificial Intelligence* (blog), *IBM*, April 24, 2024, https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning.

24. Jie Gui, Tuo Chen, Jing Zhang, et al., “A Survey on Self-Supervised Learning: Algorithms, Applications, and Future Trends,” *arXiv* (July 2024), https://doi.org/10.48550/arXiv.2301.05712.

25. Amazon, “What is RAG (Retrieval-Augmented Generation)?” *Generative AI* (blog), *Amazon Web Services*, 2025, https://aws.amazon.com/what-is/retrieval-augmented-generation/.

26. James Wang, “Scaling Is a Choice,” *Weighty Thoughts* (blog), *Substack*, December 19, 2024, https://weightythoughts.com/p/scaling-is-a-choice.

27. Intel, “Understanding Moore’s Law,” *Newsroom* (blog), *Intel*, April 9, 2025, https://newsroom.intel.com/tech101/understanding-moores-law.

28. James Wang, “Scaling Is a Choice,” *Weighty Thoughts* (blog), *Substack*, December 19, 2024, https://weightythoughts.com/p/scaling-is-a-choice.

CHAPTER 7

1. Justin Kan (@justinkan), “First time founders are obsessed with product. Second time founders are obsessed with distribution,” X, November 6, 2018, https://x.com/justinkan/status/1059989657218248704.
2. PRNewswire, “Meta Reports First Quarter 2025 Results,” news release no. 302443250, April 30, 2025, https://www.prnewswire.com/news-releases/meta-reports-first-quarter-2025-results-302443250.html.
3. Goldman Sachs, “Will the $1 Trillion of Generative AI Investment Pay Off?” *Artificial Intelligence* (blog), *Goldman Sachs Insights*, August 5, 2024, https://www.goldmansachs.com/insights/articles/will-the-1-trillion-of-generative-ai-investment-pay-off.
4. Reuters, “OpenAI does not expect to be cash-flow positive until 2029, Bloomberg News reports,” *Artificial Intelligence* (blog), *Reuters*, March 26, 2025, https://www.reuters.com/technology/artificial-intelligence/openai-does-not-expect-be-cash-flow-positive-until-2029-bloomberg-news-reports-2025-03-26/.
5. Hayden Field, “OpenAI Considering 16 States for Data Center Campuses as Part of Trump’s Stargate Project,” *AI Effect* (blog), *CNBC*, February 6, 2025, https://www.cnbc.com/2025/02/06/openai-looking-at-16-states-for-data-center-campuses-tied-to-stargate.html.
6.
    a. Casey Crownhart, “Why Microsoft Made a Deal to Help Restart Three Mile Island,” *Climate Change and Energy* (blog), *MIT Technology Review*, September 26, 2024, https://www.technologyreview.com/2024/09/26/1104516/three-mile-island-microsoft/.
    b. Brian Martucci, “Meta Seeks Up to 4 GW of New Nuclear Power to Help Meet AI, Sustainability Objectives,” *Dive Brief* (blog), *Utility Dive*, December 4, 2024, https://www.utilitydive.com/news/meta-

seeks-up-to-4-gw-of-new-nuclear-power-to-help-meet-ai-sustainability/734599/.

7. Ariana Bindman, “The Disastrous Rise and Fall of a $10 Billion Bay Area Unicorn,” *Food* (blog), *SFGate*, October 15 2024, https://www.sfgate.com/food/article/rise-fall-bay-area-startup-webvan-19829522.php.
8. Dee-Ann Durbin, “Instacart Sets IPO Price at $30 a Share, Valuing the Company at about $10 Billion,” *Business* (blog), *AP News*, September 18, 2023, https://apnews.com/article/instacart-ipo-grocery-delivery-dffca5b9d7eb9744bb565635a128f333.
9. Hayden Field, “Elon Musk’s xAI Raising up to $6 Billion to Purchase 100,000 Nvidia Chips for Memphis Data Center,” *AI Effect* (blog), *CNBC*, November 15, 2024, https://www.cnbc.com/2024/11/15/elon-musks-xai-raising-up-to-6-billion-to-purchase-100000-nvidia-chips-for-memphis-data-center.html.
10. Bernard Marr, “A Brief History of Big Data Everyone Should Read,” *Fourth Industrial Revolution* (blog), *World Economic Forum*, February 25, 2015, https://www.weforum.org/stories/2015/02/a-brief-history-of-big-data-everyone-should-read/.
11. Alex Reisner, “The Unbelievable Scale of AI’s Pirated-Books Problem,” *Technology* (blog), *The Atlantic*, March 20, 2025, https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/.
12. Rachel Metz, Shirin Ghaffary, Dina Bass, et al., “OpenAI, Google and Anthropic Are Struggling to Build More Advanced AI,” *Technology* (blog), *Bloomberg*, November 13, 2024, https://www.bloomberg.com/news/articles/2024-11-13/openai-google-and-anthropic-are-struggling-to-build-more-advanced-ai.
13. Kylie Robison, “Inside OpenAI’s $14 Million Super Bowl Debut,” *OpenAI* (blog), *The Verge*, February 9, 2025, https://www.theverge.com/openai/608476/openai-super-bowl-chatgpt-commercial.

14. Marco Quiroz-Gutierrez, “Sam Altman Says He’s Losing Money on OpenAI’s $200-Per-Month Subscriptions: ‘People Use It Much More Than We Expected,’” *Fortune* (blog), *Yahoo Finance*, January 7, 2025, https://finance.yahoo.com/news/sam-altman-says-losing-money-080700756.html.

15. OpenAI, “Hello GPT-4o,” *Milestones* (blog), *OpenAI*, May 13, 2024 https://openai.com/index/hello-gpt-4o/.

16. Miriam Reisman, “EHRs: The Challenge of Making Electronic Data Usable and Interoperable,” *Pharmacy & Therapeutics* 42, no. 9 (September 2017): 572–575, https://pubmed.ncbi.nlm.nih.gov/28890644/.

17. Richard Pallardy, “Electronic Health Record Errors Are a Serious Problem,” *Data Management* (blog), *InformationWeek*, July 9, 2024, https://www.informationweek.com/data-management/electronic-health-record-errors-are-a-serious-problem.

18. National Academies of Sciences, Engineering, and Medicine, *Artificial Intelligence and Automated Laboratories for Biotechnology: Leveraging Opportunities and Mitigating Risks* (Washington, DC: The National Academies Press, 2024), https://doi.org/10.17226/27469.

19. Aylin Sertkaya, Trinidad Beleche, Amber Jessup, et al., “Costs of Drug Development and Research and Development Intensity in the US, 2000–2018,” *JAMA Network Open* 7, no. 6 (June 2024), https://doi.org/10.1001/jamanetworkopen.2024.15445.

20. Dorothy Neufeld, “Visualizing the Training Costs of AI Models Over Time,” *Technology* (blog), *Visual Capitalist*, 2025, https://www.visualcapitalist.com/training-costs-of-ai-models-over-time/.

21. James Wang, “Does DeepSeek Wiping Out $1T of Market Value Make Sense?” *Weighty Thoughts* (blog), *Substack*, February 2, 2025, https://weightythoughts.com/p/does-deepseek-wiping-out-1t-of-market.

22. Dylan Patel and Afzal Ahmad, “The Inference Cost of Search Disruption —Large Language Model Cost Analysis,” SemiAnalysis, last modified

February 9, 2023, https://semianalysis.com/2023/02/09/the-inference-cost-of-search-disruption/.

23. Waskip, “Ask HN: How much to charge for an API call?” Hacker News, March 12, 2024, accessed May 15, 2025, https://news.ycombinator.com/item?id=39679574.

24. James Wang, “Let’s Talk About AI Power Costs,” *Weighty Thoughts* (blog), *Substack*, August 6, 2024, https://weightythoughts.com/p/lets-talk-about-ai-power-costs.

25.

    a. Casey Crownhart, “Why Microsoft Made a Deal to Help Restart Three Mile Island,” *Climate Change and Energy* (blog), *MIT Technology Review*, September 26, 2024, https://www.technologyreview.com/2024/09/26/1104516/three-mile-island-microsoft/.

    b. Brian Martucci, “Meta Seeks Up to 4 GW of New Nuclear Power to Help Meet AI, Sustainability Objectives,” *Dive Brief* (blog), *Utility Dive*, December 4, 2024, https://www.utilitydive.com/news/meta-seeks-up-to-4-gw-of-new-nuclear-power-to-help-meet-ai-sustainability/734599/.

26. Dorothy Neufeld, “Visualizing the Training Costs of AI Models Over Time,” *Technology* (blog), *Visual Capitalist*, 2025, https://www.visualcapitalist.com/training-costs-of-ai-models-over-time/.

27. Guido Appenzeller, “Welcome to LLMflation—LLM Inference Cost Is Going Down Fast,” *a16z* (blog), December 12, 2024, https://a16z.com/llmflation-llm-inference-cost/.

28. Ark Investment Management LLC, “Big Ideas 2024,” *Ark Invest* (blog), January 31, 2024, https://www.ark-invest.com/big-ideas-2024.

29. Brian Potter, “How to Build a $20 Billion Semiconductor Fab,” *Construction Physics* (blog), *Substack*, May 3, 2024, https://www.construction-physics.com/p/how-to-build-a-20-billion-semiconductor.

30. Barry Posner and Farid Tayari, “Natural Monopoly,” *Introduction to Energy and Earth Science Economics* (blog), *Pennsylvania State University*, 2023, https://www.e-education.psu.edu/ebf200/node/139.

31. Ben Thompson, “Aggregation Theory,” *Stratechery* (blog), 2025, accessed June 27, 2025, https://stratechery.com/aggregation-theory/.

CHAPTER 8

1. Intel, “Understanding Moore’s Law,” *Manufacturing* (blog), *Newsroom*, April 9, 2025, https://newsroom.intel.com/tech101/understanding-moores-law.

2. Bruce Sterling, “Real Men Have Fabs,” *Technology* (blog), *Wired*, October 9, 2008, https://www.wired.com/2008/10/real-men-have-f/.

3. Brian Potter, “How to Build a $20 Billion Semiconductor Fab,” *Construction Physics* (blog), *Substack*, May 3, 2024, https://www.construction-physics.com/p/how-to-build-a-20-billion-semiconductor.

4. Marian Tupy and Ronald Bailey, “Vastly Cheaper Computation,” *Trends* (blog), *Human Progress*, March 1, 2023, https://humanprogress.org/trends/vastly-cheaper-computation/.

5. Douglas Herz, “A Century of Moore’s Law,” SemiAnalysis, last modified February 4, 2023, https://semianalysis.com/2023/02/04/a-century-of-moores-law/.

6. Malba Tahan, *The Man Who Counted: A Collection of Mathematical Adventures* (New York: W.W. Norton & Co., 1993), 112–115.

7. Leanne Mitton, “Amdahl’s Law: Understanding the Basics,” *Learn* (blog), *Splunk*, July 27, 2023, https://www.splunk.com/en_us/blog/learn/amdahls-law.html.

8. AMD, “AMD Ryzen™ Processors for Desktops,” Ryzen Processors, AMD, last modified 2025, accessed May 9, 2025, https://www.amd.com/en/products/processors/desktops/ryzen.html.

9. Nvidia, “Compare GeForce Graphics Cards,” GeForce, Nvidia, last modified 2025, accessed May 9, 2025, https://www.nvidia.com/en-us/geforce/graphics-cards/compare/.

10. Tae Kim, *The Nvidia Way: Jensen Huang and the Making of a Tech Giant* (New York: W.W. Norton & Co., 2024), 138–140.

11. Erin Griffith, “Hardware Is the New Software? Nope, Says Marc Andreessen ‘It’s Called Hardware for a Reason—It Is Hard,’” *PandoDigest* (blog), *PandoDaily*, October 3, 2013, https://web.archive.org/web/20131004045241/http://pandodaily.com/2013/10/03/hardware-is-the-new-software-nope-says-marc-andreessen-its-called-hardware-for-a-reason-it-is-hard/.

12. Tae Kim, *The Nvidia Way: Jensen Huang and the Making of a Tech Giant* (New York: W.W. Norton & Co., 2024), 140.

13. Ibid, 144–156.

14. Luiz André Barroso, Jimmy Clidaras, and Urs Hölzle, *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*, 2nd ed. (San Rafael, CA: Morgan & Claypool, 2013): 2–6.

15. Ibid.

16. Luiz André Barroso, Jimmy Clidaras, and Urs Hölzle, *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*, 2nd ed. (San Rafael, CA: Morgan & Claypool, 2013): 20–53, 158–159.

17. Dylan Patel, “Google AI Infrastructure Supremacy: Systems Matter More Than Microarchitecture,” *SemiAnalysis* (blog), April 12, 2023, https://semianalysis.com/2023/04/12/google-ai-infrastructure-supremacy/.

18.
    a. Microsoft, “Microsoft and OpenAI Extend Partnership,” *Microsoft Corporate Blogs* (blog), January 23, 2023, https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/.

b. Amazon Staff, “Amazon and Anthropic Deepen Strategic Collaboration,” *News* (blog), *Amazon*, November 22, 2024, https://www.aboutamazon.com/news/aws/amazon-invests-additional-4-billion-anthropic-ai/.

19. Kompas AI, “The Critical Role of High Bandwidth Memory (HBM) in Modern GPUs,” *ByteBridge* (blog), *Medium*, January 9, 2025, https://bytebridge.medium.com/the-critical-role-of-high-bandwidth-memory-hbm-in-modern-gpus-63b0164f29bc/.

20. Brian Martucci, “Meta Seeks up to 4 GW of New Nuclear Power to Help Meet AI, Sustainability Objectives,” *Dive Brief* (blog), *Utility Dive*, December 4, 2024, https://www.utilitydive.com/news/meta-seeks-up-to-4-gw-of-new-nuclear-power-to-help-meet-ai-sustainability/734599/.

21. J.M.K.C. Donev, “Baseload power,” *Energy Education* (blog), *University of Calgary*, 2024, https://web.archive.org/web/20250705163025/http://energyeducation.ca/encyclopedia/Baseload_power/.

22. Casey Crownhart, “Why Microsoft Made a Deal to Help Restart Three Mile Island,” *Climate Change and Energy* (blog), *MIT Technology Review*, September 26, 2024, https://www.technologyreview.com/2024/09/26/1104516/three-mile-island-microsoft/.

23. Michael Spencer and Grace Shao, “What Are the Actual Capabilities of Huawei’s Chips and Software Systems,” *AI Supremacy* (blog), *Substack*, May 22, 2025, https://www.ai-supremacy.com/p/what-are-the-actual-capabilities.

CHAPTER 9

1. Shashank Joshi, “The War in Ukraine Shows How Technology Is Changing the Battlefield,” *Special Report* (blog), *The Economist*, July 3, 2023, https://www.economist.com/special-report/2023/07/03/the-war-in-ukraine-shows-how-technology-is-changing-the-battlefield.

2. Air Force, “B-52H Stratofortress,” *Fact Sheets* (blog), *Defense Media Activity*, June 2019, https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104465/b-52h-stratofortress/.
3. Ben Thompson, “An Interview with Gregory Allen about Transforming US Defense,” Stratechery, released October 3, 2024, 01:12:14, https://stratechery.com/2024/an-interview-with-gregory-allen-about-transforming-u-s-defense/.
4. Ibid.
5. Bob Work, “Remarks by Deputy Secretary Work on Third Offset Strategy,” *Speech* (blog), *United States Department of Defense*, April 28, 2016, https://www.defense.gov/News/Speeches/Speech/Article/753482/remarks-by-deputy-secretary-work-on-third-offset-strategy/.
6. Michael Dahm, “China’s Desert Storm Education,” *Proceedings* (blog), *US Naval Institute*, March 2021, https://www.usni.org/magazines/proceedings/2021/march/chinas-desert-storm-education.
7. Adam Bry, Ryan Tseng and Matt Cronin, “The Future of Drone Warfare,” a16z Podcast, released April 9, 2025, 00:40:31, https://open.spotify.com/episode/28Jdf8TNkGrxPPHQri3iem.
8. China Central Television, “Chinese Team Controls over 10,000 Drones with a Single Computer in a Dazzling Display,” CCTV Video News Agency, January 2, 2025, 00:02:59, https://www.youtube.com/watch?v=KxFR5zVNIqY.
9. Roger Connor, “The Predator, a Drone that Transformed Military Combat,” *Air and Space Photos* (blog), *Smithsonian National Air and Space Museum*, March 9, 2018, https://airandspace.si.edu/stories/editorial/predator-drone-transformed-military-combat.
10. Antonio Graceffo, “From Ukraine to Myanmar, Drone Warfare Marks a Paradigm Shift,” *Situation Reports* (blog), *Geopolitical Monitor*,

September 11, 2024, https://www.geopoliticalmonitor.com/from-ukraine-to-myanmar-drone-warfare-marks-a-paradigm-shift/.

11. Sinéad Baker, “Ukraine Says the Drones That Hit Russian Aircraft Used AI to Find and Strike Their Targets When They Lost Signal,” *Military & Defense* (blog), *Business Insider*, June 4, 2025, https://www.businessinsider.com/operation-spiderweb-ukraine-drones-ai-hit-russia-planes-lost-signal-2025-6.

12.

    a. James Wang, “AI Is Overturning the Military World Order,” *Weighty Thoughts* (blog), *Substack*, June 4, 2025, https://weightythoughts.com/p/ai-is-overturning-the-military-world.

    b. Military Watch Magazine Editorial Staff, “Russia’s Tu-95 Bomber Fleet Will Never Recover from Ukraine’s Latest Drone Attack: How Can the Air Force Respond?” *Force Index* (blog), June 2, 2025, https://militarywatchmagazine.com/article/russia-tu95-fleet-never-recover-ukraine-drone-attack.

13. Noah Smith, “How Chinese Drones Could Defeat America,” *Noahpinion* (blog), June 3, 2025, https://www.noahpinion.blog/p/how-chinese-drones-could-defeat-america.

14. James Wang, “AI Is Overturning the Military World Order,” *Weighty Thoughts* (blog), *Substack*, June 4, 2025, https://weightythoughts.com/p/ai-is-overturning-the-military-world.

15.

    a. Xi Jinping, “Transcript: President Xi Jinping’s Report to China’s 2022 Party Congress,” *China’s Party Congress* (blog), *NikkeiAsia*, October 18, 2022, https://asia.nikkei.com/Politics/China-s-party-congress/Transcript-President-Xi-Jinping-s-report-to-China-s-2022-party-congress.

    b. National Security Commission on Artificial Intelligence, *Final Report* (Washington, DC: National Security Commission on Artificial Intelligence, n.d.), accessed September 18, 2024, https://reports.nscai.gov/final-report/.

16. Imperial War Museums Staff, “Why Did Japan Attack Pearl Harbor,” *Stories* (blog), *Imperial War Museums*, 2025, https://www.iwm.org.uk/history/why-did-japan-attack-pearl-harbor.

17. Chad Bown, “How Trump’s Export Curbs on Semiconductors and Equipment Hurt the US Technology Sector,” *Trade and Investment Policy Watch* (blog), *Peterson Institute for International Economics*, September 28, 2020, https://www.piie.com/blogs/trade-and-investment-policy-watch/how-trumps-export-curbs-semiconductors-and-equipment-hurt.

18. Lily Ottinger and Jordan Schneider, “Emergency Edition: AI Diffusion Export Controls,” *ChinaTalk* (blog), *Substack*, January 15, 2025, https://www.chinatalk.media/p/emergency-edition-ai-diffusion-export.

19. AP Staff, “Trump Administration Rescinds Curbs on AI Chip Exports to Foreign Markets,” *Business* (blog), *Associated Press*, May 14, 2025, https://apnews.com/article/trump-biden-ai-chip-export-curbs-rescinded-bb05a9760abb8a320a447f58599e2ab6.

20. Jacques deLisle, “The Enduring—If Troubled—Genius of the Taiwan Relations Act, 45 Years On,” *Global Taiwan Brief* (blog), *Global Taiwan Institute*, April 3, 2024, https://globaltaiwan.org/2024/04/the-enduringif-troubledgenius-of-the-taiwan-relations-act-45-years-on/.

21. Emily Wu, “‘Silicon Shield’: Looking beyond Semiconductors,” *USIP Reports*, *United States Institute of Peace,* released January 2024, 00:27:24, https://web.archive.org/web/20241006035124/https://www.usip.org/publications/2024/01/silicon-shield-looking-beyond-semiconductors.

22. Viola Zhou, “TSMC’s Debacle in the American Desert,” *Features* (blog), *Rest of World*, April 23, 2024, https://restofworld.org/2024/tsmc-arizona-expansion/.

23. Taiwan Semiconductor Manufacturing Company Ltd. (TSMC), “TSMC Completed ‘The Analysis of the Effect of High Speed Rail Vibration on IC Fabrication.’ Study confirmed viability of TSMC investments, Fab sites in Tainan Science Park,” October 25, 2001, https://pr.tsmc.com/english/news/2386.

24. Steve Clemons, “The US Would Destroy Taiwan’s Chip Plants if China Invades, Says Former Trump Official,” *Email Briefings* (blog), *Semafor*, March 13, 2023, https://www.semafor.com/article/03/13/2023/the-us-would-destroy-taiwans-chip-plants-if-china-invades-says-former-trump-official.

25. Lily Ottinger and Jordan Schneider, “Emergency Edition: AI Diffusion Export Controls,” *ChinaTalk* (blog), *Substack*, January 15, 2025, https://www.chinatalk.media/p/emergency-edition-ai-diffusion-export.

26. Andrew Sharp and Ben Thompson, “The US Partners with the Middle East in AI, Why OpenAI Is Acquiring Windsurf, Google’s Side of the Platform Wars,” Sharp China, released May 19, 2025, 00:54:52, https://sharptech.fm/member/episode/the-u-s-partners-with-the-middle-east-in-ai-why-open-ai-is-acquiring-windsurf-googles-side-of-the-platform-wars.

27. Tim Kelly and Miho Uranaka, “Japan Restricts Chipmaking Equipment Exports as It Aligns with US China Curbs,” *Technology* (blog), *Reuters*, March 31, 2023, https://www.reuters.com/technology/japan-restrict-chipmaking-equipment-exports-aligning-it-with-us-china-curbs-2023-03-31/.

28. Charlotte Trueman, “Dutch Gov’t to Stop ASML from Maintaining Lithography Gear Bought by Chinese Customers,” *The Compute, Storage & Networking Channel* (blog), *Data Center Dynamics*, August 30, 2024, https://www.datacenterdynamics.com/en/news/dutch-govt-to-stop-asml-from-maintaining-lithography-gear-bought-by-chinese-customers/.

29. Alex Webb and Ian King, “The US Had a Chance to Lead in Chipmaking Tech, and Missed It,” *Technology* (blog), *Bloomberg*, April 11, 2024, https://www.bloomberg.com/news/articles/2024-04-11/euv-chipmaking-how-us-lost-control-of-cutting-edge-semiconductor-tech.

30. Norman Angell, *The Great Illusion: A Study of the Relation of Military Power to National Advantage* (Alexandria, VA: Library of Alexandria, 2007).

CHAPTER 10

1. Jonathan Yerushalmy, "'I Want to Destroy Whatever I Want': Bing's AI Chatbot Unsettles US Reporter," *Artificial Intelligence* (blog), *The Guardian*, February 17, 2023, https://www.theguardian.com/technology/2023/feb/17/i-want-to-destroy-whatever-i-want-bings-ai-chatbot-unsettles-us-reporter/.
2. Ibid.
3. Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies,* (Oxford, UK: Oxford University Press, 2014): 123–138, 150–151.
4. James Vincent, "A Video Game-Playing AI Beat Q*bert in a Way No One's Ever Seen Before," *TL;DR* (blog), *The Verge*, February 28, 2018, https://www.theverge.com/tldr/2018/2/28/17062338/ai-agent-atari-q-bert-cracked-bug-cheat/.
5. David Jang, "StarCraft 2 Protoss Legend, PartinG's Last Farewell," *StarCraft II: Legacy of the Void* (blog), *InvenGlobal*, January 5, 2022, https://www.invenglobal.com/articles/16029/starcraft-2-protoss-legend-partings-last-farewell/.
6. Alex Irpan, "An Overdue Post on AlphaStar, Part 1," *Alex Irpan* (blog), February 22, 2019, https://web.archive.org/web/20230207150338/https://www.alexirpan.com/2019/02/22/alphastar.html/.
7. Yekun Chai, "Deciphering AlphaStar on StarCraft II," *The Gradient* (blog), July 21, 2019, https://cyk1337.github.io/notes/2019/07/21/RL/DRL/Decipher-AlphaStar-on-StarCraft-II/.
8. Edward Targett, "OpenAI's Unripe 'Strawberry' Model Hacked Its Testing Infrastructure," *OpenAI* (blog), *The Stack*, September 13, 2024, https://www.thestack.technology/openais-unripe-strawberry-model-hacked-its-testing-infrastructure/.
9. Kim Martineau, "What Is AI Alignment?" *Explainer* (blog), *IBM*, November 8, 2023, https://research.ibm.com/blog/what-is-alignment-ai/.

10. Eliezer Yudkowsky, “Hard Takeoff,” *LessWrong* (blog), December 2, 2008, https://www.lesswrong.com/posts/tjH8XPxAnr6JRbh7k/hard-takeoff/.

11. Tim Mucci, “What Is the Technological Singularity?” *Think* (blog), *IBM*, June 7, 2024, https://www.ibm.com/think/topics/technological-singularity/.

12. Marc Andreessen and Ben Horowitz, “The State of AI with Marc & Ben,” *a16z Podcast*, released June 14, 2024, 01:19:49, https://a16z.com/podcast/the-state-of-ai-with-marc-ben/.

13. Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies,* (Oxford, UK: Oxford University Press, 2014): 129–130.

14. Marc Andreessen and Ben Horowitz, “The State of AI with Marc & Ben,” *a16z Podcast*, released June 14, 2024, 01:19:49, https://a16z.com/podcast/the-state-of-ai-with-marc-ben/.

15.

    a. Ken Dilanian, “Russian Trolls Who Interfered in 2016 US Election Also Made Ad Money, Report Says,” *National Security* (blog), *NBC News*, June 5, 2019, https://www.nbcnews.com/politics/national-security/russian-trolls-who-interfered-2016-u-s-election-also-made-n1013811/.

    b. Craig Silverman and Jeff Kao, “Infamous Russian Troll Farm Appears to Be Source of Anti-Ukraine Propaganda,” *Article* (blog), *ProPublica*, March 11, 2022, https://www.propublica.org/article/infamous-russian-troll-farm-appears-to-be-source-of-anti-ukraine-propaganda/.

16. David Bandurski, “China and Russia Are Joining Forces to Spread Disinformation,” *Commentary* (blog), *Brookings Institute*, March 11, 2022, https://www.brookings.edu/articles/china-and-russia-are-joining-forces-to-spread-disinformation/.

17. Center for Strategic and International Studies *(*CSIS), “Significant Cyber Incidents,” *Economic Security and Technology* (blog), *CSIS*, 2025,

accessed May 24, 2025, https://www.csis.org/programs/strategic-technologies-program/significant-cyber-incidents/.

18. Microsoft Staff, “What Is AI for Cybersecurity?” *Security 101* (blog), *Microsoft Security*, 2025, accessed May 24, 2025, https://www.microsoft.com/en-us/security/business/security-101/what-is-ai-for-cybersecurity/.

19. dkn775, “Ask HN: Anyone Else Frustrated with ChatGPT’s filters?” Hacker News, February 9, 2023, https://news.ycombinator.com/item?id=34719835.

20. Devansh, “How to Automatically Jailbreak OpenAI’s o1 [Case Studies],” *Artificial Intelligence Made Simple* (blog), *Substack*, October 12, 2024, https://artificialintelligencemadesimple.substack.com/p/how-to-automatically-jailbreak-openais/.

21. Maxime Labonne, “Uncensor Any LLM with Abliteration,” *Community Article* (blog), *Hugging Face*, June 13, 2024, https://huggingface.co/blog/mlabonne/abliteration/.

22. Leonard Lin, “An Analysis of Chinese LLM Censorship and Bias with Qwen 2 Instruct,” *Community Article* (blog), *Hugging Face*, June 11, 2024, https://huggingface.co/blog/leonardlin/chinese-llm-censorship-analysis/.

23. Erin Winick, “Amazon Ditched AI Recruitment Software Because It Was Biased against Women,” *Silicon Valley* (blog), *MIT Technology Review*, October 10, 2018, https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women/.

24. Mark Weber, Mikhail Yurochkin, Sherif Botros, et al., “Black Loans Matter: Fighting Bias for AI Fairness in Lending,” *AI Fairness* (blog), *MIT-IBM Watson AI Lab*, November 27, 2020, https://mitibmwatsonailab.mit.edu/research/blog/black-loans-matter-fighting-bias-for-ai-fairness-in-lending/.

25. Joseph Choi, “ChatGPT incorrectly diagnosed more than 8 in 10 pediatric case studies, research finds,” *Health Care* (blog), *The Hill*,

January 1, 2024, https://thehill.com/policy/health care/4387138-chatgpt-incorrectly-diagnosed-more-than-8-in-10-pediatric-case-studies-research-finds/.

26. Jack Kelly, “Google’s AI Recommended Adding Glue to Pizza and Other Misinformation—What Caused the Viral Blunders?” *Careers* (blog), *Forbes*, May 31, 2024, https://www.forbes.com/sites/jackkelly/2024/05/31/google-ai-glue-to-pizza-viral-blunders/.

27. Sara Merken, “AI ‘Hallucinations’ in Court Papers Spell Trouble for Lawyers,” *Artificial Intelligence* (blog), *Reuters*, February 18, 2025, https://www.reuters.com/technology/artificial-intelligence/ai-hallucinations-court-papers-spell-trouble-lawyers-2025-02-18/.

28. Jack Kelly, “Google’s AI Recommended Adding Glue to Pizza and Other Misinformation—What Caused the Viral Blunders?” *Careers* (blog), *Forbes*, May 31, 2024, https://www.forbes.com/sites/jackkelly/2024/05/31/google-ai-glue-to-pizza-viral-blunders/.

CHAPTER 11

1. Joel Carnegie, “‘Death of an Author: The World’s First AI-Generated Novel’ with Stephen Marche,” The Storymakers Institute, released December 5, 2023, 00:27:27, https://thestorymakersinstitute.substack.com/p/death-of-an-author-with-stephen-marche/.
2. Ibid.
3. Laura Miller, “The Robot Did It,” *Books* (blog), *Slate*, April 24, 2023, https://slate.com/culture/2023/04/ai-chatgpt-mystery-novel-death-author-stephen-marche.html.
4. Aaron Hertzmann, “How Photography Became an Art Form,” *Aaron Herztmann’s blog* (blog), August 29, 2022, https://aaronhertzmann.com/2022/08/29/photography-history.html.
5.

a. Rachel Hall, "'Dangerous Nonsense': AI-Authored Books about ADHD for Sale on Amazon," *Artificial Intelligence* (blog), *The Guardian*, May 4, 2025, https://www.theguardian.com/technology/2025/may/04/dangerous-nonsense-ai-authored-books-about-adhd-for-sale-on-amazon/.
b. Andrew Limbong, "Authors Push Back on the Growing Number of AI 'Scam' Books on Amazon," *Books* (blog), *NPR*, March 13, 2024, https://www.npr.org/2024/03/13/1237888126/growing-number-ai-scam-books-amazon/.

6. Sonja Anderson, "ChatGPT Helped Write This Award-Winning Japanese Novel," *Art Meets Science* (blog), *Smithsonian Magazine*, January 24, 2024, https://www.smithsonianmag.com/smart-news/this-award-winning-japanese-novel-was-written-partly-by-chatgpt-180983641/.
7. Anna Lagos, "A Philosopher Released an Acclaimed Book about Digital Manipulation. The Author Ended Up Being AI," *Culture* (blog), *Wired*, April 28, 2025, https://www.wired.com/story/an-acclaimed-book-about-digital-manipulation-was-actually-written-by-ai/.
8. Nick Mattiske, "He Didn't Paint His Own Paintings," *Read This* (blog), *Insights Magazine*, June 14, 2024, https://www.insights.uca.org.au/he-didnt-paint-his-own-paintings/.
9. Gary Marcus, "AI Has (Sort of) Passed the Turing Test; Here's Why That Hardly Matters," *Marcus on AI* (blog), *Substack*, April 2, 2025, https://garymarcus.substack.com/p/ai-has-sort-of-passed-the-turing.
10. Greg Kamradi, "ARC-AGI-2 + ARC Prize 2025 Is Live!" *ARC Prize* (blog), March 24, 2025, https://arcprize.org/blog/announcing-arc-agi-2-and-arc-prize-2025.
11. Aaron Kozbelt, Ronald Beghetto, and Mark Runco, "Theories of Creativity," in *The Cambridge Handbook of Creativity*, ed. James Kaufman and Robert Sternberg (Cambridge, UK: Cambridge University Press, 2010), 20–47.
12. Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (New York: Basic Books, 2018), 27.

13. Ibid, 17, 363.
14. Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (New York: Basic Books, 2018), 10, 27, 362.
15. Ibid, 349.
16. Matt Marshall, "Vinod Khosla Says Technology Will Replace 80 Percent of Doctors—Sparks Indignation," *Business* (blog), *VentureBeat*, September 2, 2012, https://venturebeat.com/business/vinod-khosla-says-technology-will-replace-80-percent-of-doctors-sparks-indignation/.
17. Siddhartha Mukherjee, "AI versus MD," *Annals of Medicine* (blog), *The New Yorker*, March 27, 2017, https://www.newyorker.com/magazine/2017/04/03/ai-versus-md.
18.
    a. Miriam Reisman, "EHRs: The Challenge of Making Electronic Data Usable and Interoperable," *Pharmacy & Therapeutics* 42, no. 9 (September 2017): 572–575, https://pubmed.ncbi.nlm.nih.gov/28890644/.
    b. Richard Pallardy, "Electronic Health Record Errors Are a Serious Problem," *Data Management* (blog), *InformationWeek*, July 9, 2024, https://www.informationweek.com/data-management/electronic-health-record-errors-are-a-serious-problem.
19. Laura Miller, "The Robot Did It," *Books* (blog), *Slate*, April 24, 2023, https://slate.com/culture/2023/04/ai-chatgpt-mystery-novel-death-author-stephen-marche.html.

CHAPTER 12

1. Robert Allen, "The Hand-Loom Weaver and the Power Loom: A Schumpeterian Perspective," (Working Paper, New York University Abu Dhabi, MAY 2017), https://nyuad.nyu.edu/content/dam/nyuad/academics/divisions/social-science/working-papers/2017/0004.pdf.
2.

a. University of Nottingham Staff, “Money,” *Weights and Measures* (blog), *University of Nottingham*, n.d., accessed May 27, 2025, https://www.nottingham.ac.uk/manuscriptsandspecialcollections/researchguidance/weightsandmeasures/money.aspx.

b. Eablanchette.com, “WORKERS: The long agony of the handloom weaver,” n.d., accessed May 27, 2025, https://web.archive.org/web/20231202000612/http://www.eablanchette.com/_supportdocs/Workers%20-%20handloom%20weavers.htm.

3. Daron Acemoglu and Simon Johnson, “Learning from Ricardo and Thompson: Machinery and Labor in the Early Industrial Revolution, and in the Age of AI” (Working paper 32416, Cambridge, MA, May 2024): 30, https://doi.org/10.3386/w32416.

4.

a. University of Nottingham Staff, “Money,” *Weights and Measures* (blog), *University of Nottingham*, n.d., accessed May 27, 2025, https://www.nottingham.ac.uk/manuscriptsandspecialcollections/researchguidance/weightsandmeasures/money.aspx.

b. Daron Acemoglu and Simon Johnson, “Learning from Ricardo and Thompson: Machinery and Labor in the Early Industrial Revolution, and in the Age of AI” (Working paper 32416, Cambridge, MA, May 2024): 27–31, https://doi.org/10.3386/w32416.

5. Communication Generation, “Fashion and the Luddites: How Textile Changes Sparked a Revolution and FF,” last modified 2025, accessed May 28, 2025, https://www.communication-generation.com/fashion-and-the-luddites-how-textile-changes-sparked-a-revolution/.

6.

a. Thomas-Earnshaw.com, “Women’s Role in the Industrial Revolution: From Factory Workers to Reformers,” *The Earnshaw Odyssey* (blog), March 15, 2024, https://thomas-earnshaw.com/blogs/the-earnshaw-odyssey/womens-role-in-the-industrial-revolution-from-factory-workers-to-reformers.

b. Cora Granata and Cheryl Koos, *The Human Tradition in Modern Europe 1750 to the Present* (Lanham, MD: Rowman & Littlefield Publishers, Inc., 2008): 31.

7.

h. John Simkin, “History of the Handloom,” *The Industrial Revolution* (blog), *Spartacus Educational*, 2020, accessed May 28, 2025, https://spartacus-educational.com/TEXweaving.htm.

i. Thomas-Earnshaw.com, “Women’s Role in the Industrial Revolution: From Factory Workers to Reformers,” *The Earnshaw Odyssey* (blog), March 15, 2024, https://thomas-earnshaw.com/blogs/the-earnshaw-odyssey/womens-role-in-the-industrial-revolution-from-factory-workers-to-reformers.

8. Marc Andreessen, “Why Software Is Eating the World,” *Andreessen Horowitz* (blog), August 20, 2011, https://a16z.com/why-software-is-eating-the-world/.

9. Simon Willison, “Not All AI-assisted Programming Is Vibe Coding (But Vibe Coding Rocks),” *Simon Willison’s Weblog* (blog), March 19, 2025, https://simonwillison.net/2025/Mar/19/vibe-coding/.

10. James Wang, “Software Engineering Is Doomed,” *Weighty Thoughts* (blog), *Substack*, July 1, 2024, https://weightythoughts.com/p/software-engineering-is-doomed.

11. StayOnTarget, “Who Are We Quoting When We Note That ‘Code Is Written Once but Read Many Times’? [closed],” *Newest Questions* (blog), *Stack Exchange*, July 24, 2020, https://retrocomputing.stackexchange.com/questions/15724/who-are-we-quoting-when-we-note-that-code-is-written-once-but-read-many-times/.

12. femio, “AI Tools Are Ironically Way More Useful for Experienced Devs Than Novices,” Reddit, April 15, 2025, https://www.reddit.com/r/ExperiencedDevs/comments/1jzpzkm/ai_tools_are_ironically_way_more_useful_for/.

13. Addy Osmani, “AI Won’t Kill Junior Devs—But Your Hiring Strategy Might,” *Elevate* (blog), *Substack*, May 10, 2025,

https://addyo.substack.com/p/ai-wont-kill-junior-devs-but-your/.

14. Olivia Hebert, “Business Insider Goes ‘All-In on AI,’ Laying off 21% of Staff,” *Bay Area & State* (blog), *SFGate*, May 30, 2025, https://www.sfgate.com/bayarea/article/business-insider-ai-laying-off-staff-20353086.php.

15. Sean O’Hare, “Automatic Translation vs Machine Translation Explained,” *Website translation* (blog), *Weglot*, February 11, 2025, https://www.weglot.com/blog/automatic-translation-vs-machine-translation-explained.

16. 데보라, “Reviewing Stitch by Google—from a Designer,” *syldeborah* (blog), *Medium*, May 26, 2025, https://syldeborah.medium.com/reviewing-stitch-by-google-from-a-designer-490fdbb84c2c.

17. @realEstateTrent, “The Leap I Witnessed in AI This Week Has Left Me Shaken,” X, March 30, 2025, https://x.com/realestatetrent/status/1906346504303894532.

18. Erin Ollila, “AI Content Rules: YouTube, Spotify & Audible 2025,” *Filler Words* (blog), *descript*, April 25, 2025, https://www.descript.com/blog/article/ai-content-on-youtube-spotify-audible.

19. Isaac Maw, “Time and Money: How Much Do Industrial Robots Cost?” *Automation* (blog), December 18, 2018, https://www.engineering.com/time-and-money-how-much-do-industrial-robots-cost/.

20. Mohamed Abdel-Aty, Shengxuan Ding, “A Matched Case-Control Analysis of Autonomous vs Human-Driven Vehicle Accidents,” *Nature Communications* 15, 4931 (2024), https://doi.org/10.1038/s41467-024-48526-4.

21. Mary Meeker, Jay Simons, Daegwon Chae, and Alexander Krey, “Trends—Artificial Intelligence (AI),” *Report* (blog), *Bond Capital*, May 30, 2025, https://www.bondcap.com/report/tai/.

22. Ben Thompson, “An Interview with Manna Founder and CEO Bobby Healy About Drone Delivery,” *Stratechery* (blog), February 20, 2025, https://stratechery.com/2025/an-interview-with-manna-founder-and-ceo-bobby-healy-about-drone-delivery/.

23. Alicia Kelso, “Chick-fil-A Pilots Robot Delivery as Investment in Autonomous Tech Grows,” *Dive Brief* (blog), *Restaurant Dive*, https://www.restaurantdive.com/news/chick-fil-a-pilots-robot-delivery-as-investment-in-autonomous-tech-grows/598339/.

24. Ark Investment Management LLC, “BIG IDEAS 2024,” *Ark Invest* (blog), January 31, 2024, https://www.ark-invest.com/big-ideas-2024.

25. Lindsey Wilkinson, “United Airlines CEO: ‘We’re Probably Doing More AI Than Anyone,’” *Dive Brief* (blog), *CIO Dive*, June 2, 2025, https://www.ciodive.com/news/united-airlines-ceo-AI-use-cases/749563/.

26. Nathan Lambert, “People Use AI More Than You Think,” *Interconnects* (blog), *Substack*, May 21, 2025, https://www.interconnects.ai/p/people-use-ai-more-than-you-think.

27. Olivia Moore and Marc Andrusko, “What ‘Working’ Means in the Era of AI Apps,” *a16z* (blog), *Andreessen Horowitz*, https://a16z.com/revenue-benchmarks-ai-apps/.

CHAPTER 13

1. Jake Renzella and Vlada Rozova, “The ‘Dead Internet Theory’ Makes Eerie Claims about an AI-Run Web. The Truth Is More Sinister,” *Newsroom* (blog), *University of New South Wales Sydney*, May 20, 2024, https://www.unsw.edu.au/newsroom/news/2024/05/-the-dead-internet-theory-makes-eerie-claims-about-an-ai-run-web-the-truth-is-more-sinister/.

2. Erica Pandey, “AI Cheating Surge Pushes Schools into Chaos,” *Technology* (blog), *Axios*, May 26, 2025, https://www.axios.com/2025/05/26/ai-chatgpt-cheating-college-teachers.

3. Shashi Baltutis, "'More Efficient and Easier': How AI Can Help Teachers Spend Less Time on Things Other Than Teaching" *South Australia Education* (blog), *The Advertiser*, May 27, 2025, https://www.adelaidenow.com.au/south-australia-education/more-efficient-and-easier-how-ai-can-help-teachers-spend-less-time-on-things-other-than-teaching/news-story/483005387b9aca81929c11d4376b6f31.
4.
    a. Saloni Dattani, Lucas Rodés-Guirao, and Max Roser, "Fertility Rate," Our World in Data, last accessed June 1, 2025, https://ourworldindata.org/fertility-rate.
    b. Our World in Data, "Replacement-Level Fertility Rate," Grapher, last accessed June 1, 2025, https://ourworldindata.org/grapher/replacement-fertility-rate.
5. Louis March, "To the Surprise of Demographers, African Fertility Is Falling," *Mercator* (blog), September 19, 2024, https://www.mercatornet.com/to_the_surprise_of_demographers_african_fertility_is_falling.
6. Jesús Fernández-Villaverde, "El declive demográfico de la humanidad," *Economía* (blog), *El Confidencial*, February 18, 2024, https://blogs.elconfidencial.com/economia/la-mano-visible/2024-02-18/declive-demografico-humanidad_3831566/.
7. James Connor, "Neil Howe's Dire Prediction: Economic Stagnation and the End of Prosperity," *Wealthion*, released March 7, 2024, 00:54:30, https://wealthion.com/neil-howes-dire-prediction-economic-stagnation-and-the-end-of-prosperity/.
8.
    a. World Bank, *Age Dependency Ratio: Older Dependents to Working-Age Population for the United States [SPPOPDPNDOLUSA]*, 2023, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/SPPOPDPNDOLUSA.
    b. World Bank, *Age Dependency Ratio: Older Dependents to Working-Age Population for Germany [SPPOPDPNDOLDEU]*, 2023, Federal

Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/SPPOPDPNDOLDEU.

c. World Bank, *Age Dependency Ratio: Older Dependents to Working-Age Population for China [SPPOPDPNDOLCHN]*, 2023, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/SPPOPDPNDOLCHN.

9. George Prior, "Japan's Birth Crisis Is a Leadership Failure," *Opinion* (blog), *Asia Times*, February 28, 2025, https://asiatimes.com/2025/02/japans-birth-crisis-is-a-leadership-failure/.

10. Jaehyun Eom, "Korea Says Its Pension Fund Will Run Dry by 2056 If No Reform," *Markets* (blog), *Bloomberg*, September 24, 2024, https://www.bloomberg.com/news/articles/2024-09-25/korea-says-its-pension-fund-will-run-dry-by-2056-if-no-reform/.

11. Leslie Kramer, "What Is GDP and Why Is It So Important to Economists and Investors," *Guide to Economics* (blog), *Investopedia*, September 10, 2024, https://www.investopedia.com/ask/answers/what-is-gdp-why-its-important-to-economists-investors/.

12. Nam Kyu Kim, "Revisiting Economic Shocks and Coups," *Journal of Conflict Resolution* 60, no. 1, (2016): 3–31, https://www.jstor.org/stable/24755878.

13. Diala Al Masri, Vizhdan Boranova, Steven Dang, et al., "Fragile Foundations: The Lasting Economic Scars of Conflict," *Regional Economic Outlook: Middle East and Central Asia* (blog), *International Monetary Fund*, April 2024, https://www.imf.org/-/media/Files/Publications/REO/MCD-CCA/2024/April/English/ch2.ashx.

14. Adam Volle, "Occupy Wall Street," https://www.britannica.com/topic/Occupy-Wall-Street.

15. Janet Nguyen, "Incumbents Are Losing around the World, Not Just the US," *Election 2024* (blog), *Marketplace*, November 14, 2024, https://www.marketplace.org/story/2024/11/14/incumbents-are-losing-around-the-world-not-just-the-u-s.

16. David Frankford, “The Remarkable Staying Power of ‘Death Panels,’” *Journal of Health Politics, Policy, and Law* 40, no. 5, (2015): 1,087–1,101, https://doi.org/10.1215/03616878-3161212.

17. Jonquilyn Hill, “How Gen Z Found Religion,” Explain It to Me, June 1, 2025, 00:29:16, https://open.spotify.com/episode/7LGjqL72wuw7dqNcxBEqMs.

CONCLUSION

1. @startuparchive_, “Steve Jobs Explains Why Computers Are a Bicycle for the Mind: ‘We Humans Are Tool Builders,’” Startup Archive, April 27, 2024, 00:01:12, https://www.youtube.com/watch?v=NjIhmzU0Y8Y.

2. Henry Lowood, “Doom,” *Games, Hobbies, & Recreational Activities* (blog), *Encyclopaedia Britannica*, May 22, 2025, https://www.britannica.com/topic/Doom.

3. @NextGenDreams, “Cyberpunk 2077 Goes Hyper-Realistic: POV Night Life & Showcase with DreamPunk 2.0 Graphics in 8K,” NextGen Dreams, August 16, 2024, 00:04:45, https://www.youtube.com/watch?v=90oVkISQot8.

4. Kwan Tan, “A Former Top Meta Exec Says AI Is Doomed if Artists Want Tech Firms to Ask for Permission to Train Models on Their Work,” *AI* (blog), *Business Insider*, May 27, 2025, https://www.businessinsider.com/ex-meta-exec-ai-doomed-tech-firms-ask-artists-permission-2025-5.

5. Jaspreet Singh, “*New York Times* Partners with Amazon for First AI Licensing Deal,” *Business* (blog), *Reuters*, May 29, 2025, https://www.reuters.com/business/retail-consumer/new-york-times-amazon-sign-ai-licensing-deal-2025-05-29/.

6. Nathan Lambert, “How Language Model Post-Training Is Done Today,” Interconnects AI, January 8, 2025, 00:53:50, https://www.youtube.com/watch?v=6yIMb0K-aS4.

7. Ben Thompson, “An Interview with OpenAI CEO Sam Altman about Building a Consumer Tech Company,” Stratechery, March 20, 2025,

00:58:14, https://stratechery.com/2025/an-interview-with-openai-ceo-sam-altman-about-building-a-consumer-tech-company/.

8. Nathan Lambert, “People Use AI More Than You Think,” *Interconnects* (blog), *Substack*, May 21, 2025, https://www.interconnects.ai/p/people-use-ai-more-than-you-think.
9. Alexander Atkins, “How Many Words Does the Average Person Speak in a Lifetime?” *Atkins Bookshelf* (blog), *WordPress.com*, May 7, 2014, https://atkinsbookshelf.wordpress.com/tag/how-many-words-does-the-average-person-speak-in-their-lifetime/.
10. OpenAI, “What Are Tokens and How to Count Them?” OpenAI Help, n.d., accessed June 11, 2025, https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them.
11. Nathan Lambert, “Qwen 3: The New Open Standard,” *Interconnects* (blog), *Substack*, April 28, 2025 https://www.interconnects.ai/p/qwen-3-the-new-open-standard.
12. SoA Policy Team, “SoA Survey Reveals a Third of Translators and Quarter of Illustrators Losing Work to AI,” *News* (blog), *The Society of Authors*, April 11, 2024, https://societyofauthors.org/2024/04/11/soa-survey-reveals-a-third-of-translators-and-quarter-of-illustrators-losing-work-to-ai/.
13. Devansh, “How DeepMind Built AI That Evolves Itself [Breakdowns],” *Artificial Intelligence Made Simple* (blog), *Substack*, May 16, 2025, https://artificialintelligencemadesimple.substack.com/p/how-deepmind-built-ai-that-evolves.
14. Ali Behrouz, Peilin Zhong, and Vahab Mirrokni, “Titans: Learning to Memorize at Test Time,” *arXiv*, (December 2024), https://doi.org/10.48550/arXiv.2501.00663.
15. Bo Peng, Eric Alcaide, Quentin Anthony, et al., “RWKV: Reinventing RNNs for the Transformer Era,” *arXiv*, (May 2023), https://doi.org/10.48550/arXiv.2305.13048.

16. Bob Work, “Remarks by Deputy Secretary Work on Third Offset Strategy,” *Speech* (blog), *United States Department of Defense*, April 28, 2016, https://www.defense.gov/News/Speeches/Speech/Article/753482/remarks-by-deputy-secretary-work-on-third-offset-strategy/.

17. Bill Gurley, “How to Miss by a Mile: An Alternative Look at Uber’s Potential Market Size,” *Above the Crowd* (blog), July 11, 2014, https://abovethecrowd.com/2014/07/11/how-to-miss-by-a-mile-an-alternative-look-at-ubers-potential-market-size/.

18. Ibid.

19. Google, “Uber Technologies Inc,” Google Finance, last accessed June 2, 2025, https://www.google.com/finance/quote/UBER:NYSE.

20. Bill Gurley, “How to Miss by a Mile: An Alternative Look at Uber’s Potential Market Size,” *Above the Crowd* (blog), July 11, 2014, https://abovethecrowd.com/2014/07/11/how-to-miss-by-a-mile-an-alternative-look-at-ubers-potential-market-size/.

21. Ibid.

22. Gary Gibbons, “Serendipity and the Prepared Mind: An NHLBI Intramural Researcher’s Breakthrough Observations,” *Director’s Messages* (blog), *National Heart, Lung, and Blood Institute*, December 24, 2013, https://www.nhlbi.nih.gov/directors-messages/serendipity-and-the-prepared-mind/.

This file was downloaded from Z-Library project



Your gateway to knowledge and culture. Accessible for everyone.

z-library.sk z-lib.gs z-lib.fm go-to-library.sk

Official Telegram channel

Z-Access

https://wikipedia.org/wiki/Z-Library