INTRODUCTION

Turing's Paradox *and the* Failure *of the* Sciences of Man

> The works and customs of mankind do not seem to be very
> suitable material to which to apply scientific induction.
> —Alan Turing, "Computing Machinery and Intelligence" (1950)

> We may accomplish the AI dream by stripping humans
> of so much singular identity that people are reduced to
> mere agents.
> —Illah Reza Nourbakhsh, *Robot Futures* (2015)

THE ORIGINS OF ARTIFICIAL INTELLIGENCE (AI) CAN BE FOUND IN THE history of simple people, not smart machines. This claim requires some qualification. To say that AI is the result of simple—or, more precisely, simplified—people is not to deny the incredible increase in complexity in modern thinking machines, and it certainly is not to deny the intelligence of those responsible for the algorithms, computer programs, and data collection tools behind them. Machines can certainly do more things that appear to be thought-like than they previously have been able to do, and the creators of these machines were and are unquestionably intelligent people, at least by the standards set for intelligence in most of the world in the twenty-first century. Rather, to say that simple people are responsible for AI is to argue for the historical antecedents of roboticist Illah Nourbakhsh's provocative claim that the "AI dream" may occur through "stripping humans" of their "singular identity."[1] As will be seen, in both the abstract and the concrete sense, simple people have been central to the idea of artificial intelligence for centuries.

More specifically, this book will argue that the idea of artificial intelligence emerged out of reductive and simplistic methodologies pursued in the

collective fields of knowledge generally known as the social sciences, or what were for centuries called the "sciences of man."[2] Going back four centuries to the idea of "mechanical philosophy"—which attempted to describe nature through machine analogies—the book covers a number of approaches that sought to study human thought and behavior using the assumptions, tools, and techniques of the natural sciences. This includes long-forgotten ideas like the "geometric spirit," "social physics," "the hedonistic calculus," and the "iron laws" of classical political economy, as well as more recent approaches like eugenics, statistical sociology, positivist economics, and behaviorism. Each, in their own way, attempted to reduce the wide diversity and seeming randomness of human thought and behavior into a model that could predict, often quantitatively, the actions of human subjects. In addition to presenting the abstract theories that simplified human behavior, the book also explains how the failure of such ideas to accurately predict human actions led to interventionist actions in the lives of real people. As the messy business of history frequently interfered with the attempts to simplify human behavior to lawlike consistency, social thinkers often found that such "laws" could only obtain through education, social reform, and political intervention. Finally, the book also covers the arguments of a number of *critics* of such approaches, including social thinkers, scientists, and mathematicians who worried that theories of human action driven exclusively by scientific methodologies had the potential to create, rather than reveal, a new type of human subject. In their view, reductive theories produced a very real kind of artificial intelligence, one that often seemed to originate in the practices of scientific work itself.

Remarkably, the successes and failures of the sciences of man and social sciences to understand human thought and action over nearly four hundred years have been almost completely absent in discussions of the history of artificial intelligence. In part, this oversight can be explained by Alan Turing's comment above, that science was unsuited to studying the "works and customs of mankind."[3] In what may seem paradoxical, Turing's skeptical stance on using "induction"—i.e., the process of deriving laws from observation and experiment—to study human action occurred in the same 1950 paper in which he first proposed his famous "Turing Test," a thought experiment that predicted smart machines would one day be able to fool human examiners by "imitating" a person. As a prophet of artificial intelligence, founder of computer science, and champion of the scientific method, it may seem odd that Turing doubted that science could understand people. How, after all, could scientists fashion a machine that imitated something that could not be

understood through scientific methodology? If people's actions and behavior could not be understood through "scientific induction," then how could science build machines that imitated this action? If *social scientists* could not make reliable predictions about people based on previous actions, how could *computer scientists* build something that could imitate this unpredictability? The answer, explored in this book, was to simplify what it means to be a person.

Although human thought, behavior, and communication have been the goal for AI since Turing, little has been accomplished in the subsequent seventy years that might suggest computer scientists have looked to the history of the social sciences as a guide to understanding their ultimate goal. In fact, to look at the state of artificial intelligence today is to see that questions about the nature of human thought, behavior, and communication have remained unasked, unanswered, or ignored, in large part because AI history has been so narrowly focused on machines and their makers as to bracket out discussions of what kind of *person* AI is supposed to replicate. For example, in interviews with fifty leading AI experts published in 2018, the futurist and writer Martin Ford asked each subject to provide a year when "human-level AI might be achieved."[4] While the predictions ranged from a few decades to a century, neither Ford nor any of those interviewed defined "human level" or considered how such a standard could be determined scientifically. Though no scientific account was offered of what it actually means for a person to act and think, all fifty engineers and computer scientists were confident that AI would one day reach this undefined goal.

In perhaps the most vivid illustration of how often the "human level" is usually ignored in discussions of AI, in both Turing's legendary "imitation game" and its many subsequent iterations, there are few accounts of what kind of *people* might participate alongside the machine. For example, in 2002 the futurist Ray Kurzweil and businessman Mitch Kapor made a bet about the possibility that a machine might one day pass the Turing Test. Although the "rules" of the test ran close to two thousand words, the only mention of the human players was the stipulation that three human "foils" were needed, without any mention of, say, the age, race, gender, religious background, or personality of the people selected to participate.[5] Even putting a barrier like culture aside, it could be argued that machines would have varying levels of success at the Turing Test with eccentrics, artists, mystics, or children, and it is perhaps unsurprising that the first computer to "pass" the test did so through mimicking the random interruptions, non sequiturs, and "b*llshit" of a teenager.[6] In fact, one could imagine in all seriousness that the easiest way

for a machine to pass the Turing Test would be to have either a very boring human competitor or a dumb interlocutor. Similarly, to give just one notable example of how AI success might depend upon the simplification of human activity—the Writers Guild of America strike from the summer of 2023—it has been argued that years of simplistic, "derivative," and "formulaic" script writing has made the machines' task that much easier.[7] Machine success, therefore, is not necessarily tied to a fixed goal but might be greatly helped by a lowering, or simplification, of the "human level."

The idea that new technologies might inhibit human agency and intellect is as old as warnings from Socrates and Plato that written language harmed human memory, and the perceived threats of various machines have been debated many times over.[8] Rather than adopt a simplistic technological determinism that blames machines themselves for a simplification of human thought and behavior, this book instead argues that the "human level" that AI researchers are trying to reach would be unimaginable without the long history of scientific attempts to understand human thought, action, and behavior. People have of course come up with many other methods for trying to understand one another throughout history outside of the social sciences—from divine explanation to art and literature to intuition—but these approaches are particularly unsuited as forerunners of today's AI, as they tend to make people's "singular identity" paramount. An AI system based on individual human beings as divine creations possessing willing souls, complex literary figures rent by deeply personal family histories, or completely absurd and random beings would not make it very far. Conversely, many things that AI does do well would seem completely absurd or pointless to a vast number of people in different times and places who have not relied on scientific explanations for human behavior. To a medieval European peasant, Māori oral historian, Tang dynasty chronicler, or Roman centurion, for example, there would be very little utility or intelligence in today's machines, or anything "human level" in AI success in playing games, scheduling meetings, summarizing legal texts, or producing middling undergraduate papers. What this book attempts to show, therefore, is that the multifaceted phenomenon we know today as AI has been made possible because of a significant effort over the past four hundred years to reduce the "works and customs of mankind" to a point where scientific methodologies could be used to understand them. For the harshest critics of reductive social science traced in this story, it was theories about simple people, rather than smart machines, which proved the stuff of nightmares.

The Quest

In shifting the focus of the story of artificial intelligence from technical developments in machines to scientific theories of people, this book significantly expands the chronological scope of AI history.[9] Most standard histories of the field begin approximately seventy to ninety years ago and are centered around a few crucial figures in computer science and mathematics.[10] In this narrative, the "quest" for AI begins in 1936 with the twenty-four-year-old Turing's remarkable "Entscheidungsproblem" paper, which required Turing to define the idea of a "computable number."[11] In the process of doing so, Turing also imagined a machine that could compute such numbers, which led to two groundbreaking papers that suggested such a hypothetical machine could therefore replicate the process of human computation and even thought.[12] Inspired by Turing and the growth of electronic computers, in 1955 four researchers requested $13,500 for a summer conference at Dartmouth to study "Artificial Intelligence," with the belief that any "aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." As they ambitiously declared, the Dartmouth group believed that a "significant advance" could be accomplished if a "carefully selected group of scientists work on it together for a summer."[13]

For most of early AI history, researchers were split between two approaches that Turing had suggested for building an intelligent machine. The first, supported by two of the Dartmouth Conference organizers, Marvin Minsky and John McCarthy, envisioned that machines could be programmed with sufficient rules to both process and produce strings of "symbols" that could mimic human cognition and language. While far from a complete failure—it was this kind of machine that IBM eventually used to beat Garry Kasparov at chess—"symbolic" AI largely foundered following a number of spectacular failures with expensive and clunky robots in the 1970s and 1980s. In 1973, an influential report from the United Kingdom that dismissed the idea as close to a joke almost buried symbolic AI completely.[14]

However, in his canonical papers Turing had also suggested that machine intelligence could be created by building a more basic machine that "learned" like a child. The idea was then given formal shape in a groundbreaking paper coauthored by the psychologist Warren S. McCulloch and the neuroscientist Walter Pitts in 1943, which suggested that the nervous system, and therefore the mind, might function in a similar binary way as formal mathematical logic.[15] The idea of building such a "neural network" had been included in the original Dartmouth proposal, and Frank Rosenblatt in 1958 had offered

a tantalizing glimpse of a learning machine in his "Perceptron," but the idea of building brains lost significant interest and funding in the wake of a book coauthored by Minsky in 1969 that questioned neural networks.[16] With both approaches stymied in the 1970s and 1980s—the "AI winters"—neural networks, or "connectionist AI," returned triumphantly in the 2000s, most spectacularly at the 2012 ImageNet competition, where a "neural net" built by a team from the University of Toronto trounced its symbolic AI competitors in vision recognition.[17] It is these connectionist nets, now armed with extraordinary data harvested from the internet by technology companies, that promise to transform the world for good or evil.[18]

Though artificial intelligence has produced a robust literature—from technical guides to personal memoirs to deeply philosophical work to polemics against technology—most accounts largely rely on the narrative above and describe the history of AI as a somewhat fitful but ultimately progressive story of AI approaching the human level.[19] For example, in just one of the many "quest" narratives, the "grand goal" of "human-level AI" is presented as a project that should "develop artifacts that can do most of the things that humans can do . . . specifically those things that are thought to require intelligence."[20] Versions of this claim are repeated in many books on AI, and the idea of what has historically constituted those "things that require human intelligence" is either assumed or ignored. Yet, as historians of science have shown, the "human level" is not a static category. In the late 1800s, for example, "human intelligence" would have included the ability to construct a personal calendar, do basic rote mathematics, and measure the progress of celestial objects. For most of the twentieth century, "human intelligence" would have also meant the ability to take dictation or review legal documents, tasks now taken over in large part by machines. As shown in one classic study in the history of science, even tracing the movement of microscopic particles was once the province of human intelligence rather than machines.[21] Though some of these developments were driven by smart machines, others occurred through shifts in labor practices or cultural norms. Indeed, as one historian has noted, the ability to perform advanced calculations went from being the highly intelligent "distinctive activity of a scientist" to the work of "an anonymous drudge" at just around the moment when such labor shifted from men to women.[22] To judge by this history, then, the "grand goal" of the AI quest has already been accomplished several times over, often centuries before the term was coined.

While recent histories of artificial intelligence have provided a clear and triumphant narrative, the historian of science Stephanie Dick observed in a

recent critique of histories of the field, "There isn't a straightforward narrative of artificial intelligence from the 1950s until today."[23] In her view, neural networks and generative AI—or what Dick calls "inhuman artificial intelligence"—might in fact represent a *retreat* from the original aims of Turing and the Dartmouth Conference, as the early dreams of AI and the modern realities seem related "in name only." Indeed, the proliferation of terms deployed by experts in the field—weak and strong AI, artificial general intelligence, and good old-fashioned AI—seem to indicate a splintering rather than a synthesis of ideas. While the original Dartmouth organizers had assumed AI would imitate human behavior, many modern researchers have dismissed the idea that their machines might be doing anything that resembles human thought. On the one hand, AI is often described as the same project developed at Dartmouth, for example when Eric Horvitz, director of Microsoft Research, testified to the US Congress in November 2016 that "AI is . . . aimed at a shared aspiration, the scientific understanding of thought and intelligent behavior, and in developing computing systems based on these understandings."[24] On the other hand, the limitations of neural nets as a model of human cognition have been pointed out in a series of recent articles and books asking whether the field is now a "one-trick pony" that offers flashy demonstrations but no legitimate explanation of how the human mind works.[25] As Margaret Boden claimed in 2016, representing some but certainly not all research, AI does not even try "to mimic human intelligence" anymore, opting instead for useful tools.[26] Confirming Boden's point, the enthusiasm in 2023 for generative AI further transformed the idea from a model of the human mind to simply one more profit-making exercise for Silicon Valley investors.[27] Regardless of which "trick" wins out in the contest for the dominant online chatbot, it can be argued that such confusion over AI occurs because its history has been separated from a longer history of thinking about human beings. As Dick concluded in her broadside against the quest narrative, "Artificial intelligence belongs in the history of *human intelligence*."[28]

Why a Long History of AI?

Expanding the scope of what counts in the history of AI has a number of important consequences. As seen above, the first consequence of a longer approach is to challenge the standard narrative at the heart of many philosophical and political debates about AI: that there exists a fixed standard of the "human level," which machines are approaching and will, eventually, surpass. Once restricted to dystopian novels, science fiction films, and the

Unabomber manifesto, this vision emerged in serious literature in the second decade of the twenty-first century, imagining the concept of human as indelibly fixed while machine intelligence is gradually "rising" to the point where it matches and exceeds the human.[29] Such a vision—often referred to as transhumanism—holds sway too in popular influencers in the world of computers and technology, where human life is either existentially imperiled or rapturously transformed.[30] Unsurprisingly, much of the concern and the enthusiasm has come from individuals heavily invested (in multiple senses) in machine technology and who harbor a view that historians and philosophers call "technological determinism," a vision of history where new technologies emerge out of a socioeconomic and cultural vacuum to transform unwitting and passive peoples.[31] Yet, as many of these same historians and philosophers have been at pains to point out, the determinist view misreads the historical complexity and agency of people themselves, as human goals, behavior, and intelligence have been in constant flux for centuries, just as often shaping the machines as they are shaped by them.

The second consequence for a history of AI told from the perspective of the sciences of man and social sciences is that it redirects attention to the amount of human labor needed to help build and run the machines. As the historian and sociologist of science Simon Schaffer noted, if "machines look intelligent," it is because "we do not concentrate on where their work is done."[32] Indeed, the artificial intelligence scholar Kate Crawford's recent *Atlas of AI* performs exactly this task, noting that an "abstracted analysis" of AI has blinded us to the "embodied and material" consequences of AI work.[33] As David Alan Grier noted in his survey of "human computers," during the nineteenth and early twentieth centuries, astronomers, mathematicians, surveyors, and many other "gentlemen of science" were forced to substitute "brawn for brain," using "immense labor" to complete scientific work.[34] Based on the principles of division of labor, social scientists of the nineteenth and twentieth centuries were also conscious about the need to create institutions and work practices that had proved successful in the natural sciences. As the resources and manpower devoted to astronomy, chemistry, and physics increased, so too did resources to study people using scientific tools.[35]

In spite of what has often been heralded as the creative spirit of modern technology companies, the labor practice in much AI work today too seems drawn from many aspects of "industrial science," the form of hierarchical labor first used in observatories, university laboratories, and other social scientific institutions of the past.[36] As one 2014 headline put the matter, most

of the data processing that is central to AI today requires heroic "Janitor Work" to sift, mold, and shape the data, with 50 to 80 percent of the work dedicated to the "mundane task of selecting and preparing unruly digital data."[37] The 2018 German documentary *The Cleaners* also makes it clear that warehouses full of low-paid immigrant laborers in Europe are needed to sift through the enormous amount of racism, pornography, and violence on the internet that would otherwise render most AI machines as spiteful and hateful beings.[38] As a recent spate of books have shown, a rigidly disciplined hierarchical labor of "human-fueled automation" is at the heart of the supposedly "productive, nonhierarchical, and playful workplace" described in the earliest reports from tech campuses.[39] Recent studies have also connected AI success to the rampant poverty of contributors to sites like the Mechanical Turk and LeadGenius, to say nothing of the unpaid labor of the billions of people who provide the data necessary to train AIs.[40] In Heike Geissler's novel *Seasonal Associate*, for example, the roots of success for a massive digital marketplace is found not in the algorithms themselves but rather in the processing of human labor through constant monitoring and surveillance.[41] And in celebrated memoirs of recent years, even the supposedly glamorous world of tech work becomes a slog, recalling the dim awakening of a collective conscious seen in the factories of England in the 1830s.[42] Indeed, in their custodial metaphors, recent accounts of AI labor practices give new meaning to the historian of science Thomas Kuhn's famous invocation of the "mopping up operations" at the heart of modern science.[43]

The third consequence of reframing AI as the product of a long history of social scientific attempts to reduce human thought and behavior is that it can help explain one of the most fraught aspects of AI today: bias in the machines.[44] While it is plainly obvious that AI algorithms exhibit bias against women, minorities, and the poor, a philosophy of technical determinism has left technical "fixes" from the "coding elite" as the only solution for the "invisible men and women" who help produce AI.[45] As historians have shown, however, ideologies of European cultural superiority were co-produced with machines in the age of the scientific revolution and European exploration.[46] So too have historians shown that data collection often involves the "invisible labor" of the most marginalized groups, and a long history of the social scientific roots of AI demonstrates that bias has been embedded from the beginning.[47] The book's focus on European and American thinkers, mostly men, is therefore not because this history is more interesting or because the conclusions reached by these thinkers are any more true than other attempts

to understand humanity through scientific means. Nor is it because much, though certainly not all, historical work in the social sciences and in AI has been done by men of European descent. Indeed, it might be argued that because modern AI has in recent years become so transnational and attracted so many practitioners from a variety of backgrounds, a narrow focus on Western thinkers—and even archaic terminology like "science of man"—obscures the modern diversity of the field.

Yet, as seen throughout the book, the centrality of European and American thinkers to AI and the social sciences is part of a corollary argument, developed in the final two chapters, that the concepts and codes embedded into the earliest vision of AI were grounded in Western social scientific thought.[48] For example, historians have shown that it was not a coincidence that AI and the modern social sciences emerged in America at almost the exact same time as "modernization theory," the belief that a particular Western form of politics and culture would necessarily predominate throughout the world.[49] Given that these same assumptions are under attack today, and that AI seems remarkably rife with bias for supposedly neutral machines, it is worth noting at the outset that AI may represent a last attempt to preserve a culturally specific and historically determined definition of human thought and behavior against challenges from non-Western perspectives on the meaning and purpose of human life.

Histories of People, Data, and Machines

The justification for a longer approach to AI history has been based in research and work in the history of science, a field well positioned to add context and understanding to the teleological and internal approaches that have mostly dominated and distorted AI history. Indeed, it was something of a surprise to many historians of science when a 2017 *New York Times* opinion piece criticized academia for failing to contribute to the story of modern technology, blaming the "ivory tower" for "being asleep at the wheel" and having "essentially no distinct field of academic study that takes seriously the responsibility of understanding and critiquing the role of technology."[50] For those in the ivory tower who have labored in the "distinct fields" of history of science, philosophy of science, and science and technology studies, nothing could be further from the truth. Rather, these fields have been attempting to understand technology for decades, producing nuanced histories of past machines, transformative works on quantification, and remarkably detailed histories and ethnographies of the kind of data science that make up today's AI machines.

No book has systematically attempted to connect the development of artificial intelligence to a long history of social scientific methodologies, but there is nevertheless a robust legacy of studies that makes such a book possible. Boden's *Mind as Machine* provided a vivid history of thinkers who had tried to define thinking as a mechanical process, and recent works by Matthew Jones and Jessica Riskin have also explored the considerable overlap between machines and ideas since the scientific revolution, providing richly sourced accounts of the search for agency in matter.[51] In *Sublime Dreams of Living Machines*, Minsoo Kang traveled back even further to investigate how automatons have been imagined since the days of Hesiod and Hephaestus, emphasizing how intellectual and cultural contexts can shape the creation of machines.[52] Adrienne Mayor's *Gods and Robots* too explored the importance of *ideas* in directing ancient machine-making, noting that the "black box" of ancient machines—the precise mechanisms of the bronze giant Talos, for example—might seem a handy metaphor for our own ignorance of modern proprietary algorithms and corporate technologies.[53] Even in the modern age of machines and the sciences of man, more specialized prehistories of artificial intelligence are a reminder that human cultures and ideas shape machines rather than vice versa. In Adelheid Voskuhl's *Androids in the Enlightenment* and Kevin LaGrandeur's *Androids and Intelligent Networks in Early Modern Literature*, for example, the authors examine how European machine makers and consumers valued feeling, affect, and sentimentality over intelligence.[54] In an irony for the Age of Reason, it was their very ability to *appear* artificial that made machines intelligent. The concern then was not that the machines would become intelligent but that people would come to look like automatons.

While these histories have provided key demonstrations of how culture and ideas can shape machines, they have not been primarily concerned with the role of the social sciences in shaping the kind of "human" to which the machines might aspire. In contrast, the story that follows draws on a larger context of thinking about data and social science that has emerged in recent decades, dating back at least to a conference and 1981 book on the "Probabilistic Revolution."[55] In this work, historians of sciences were particularly attentive to how governments, social scientists, and technology companies have used data collection to reduce the broad and messy idea of a person to a scientifically intelligible concept. In the 1980s and 1990s, a number of groundbreaking studies emerged where individual intelligence was redefined as "calculation" and where people and societies were redefined as quantifiable entities.[56] Outside of these more specialized books, other works on quantification demonstrated how

numbers became associated with objectivity, which has subsequently shaped social scientific research.[57] Decades later, the literature on the quantitative revolution has moved into popular business and trade books, as the "boring but necessary" world of data exploded onto the bestseller lists in the twenty-first century, providing a mantra for the fusion of data and capital.[58] As an indication that much work remains to be done, however, a recent popular book on data has been praised for its "mind-altering insight . . . that the numbers we use to capture the human experience are themselves a form of creative story-telling."[59] For those familiar with the works described above, such "mind-altering insights" have been available for decades.

While the major revolutionary studies of quantitative history are close to thirty to forty years old, appearing before machine personal assistants, self-driving cars, and predicative algorithms became widespread sources of excitement and fear, a newer body of literature has begun to emerge that explores the recent impact on human intelligence and behavior resulting from the reductive process of quantification in the social sciences. In these works, scholars show how recent data collection has morphed from the days of crude averages and bell curves, as American corporations have tried to create individualized quantitative "selves" for the purpose of manipulation.[60] This new "turn" in data and quantitative histories can also been seen in two recent journals dedicated to the history of data collection processes. In one series of articles on the "Histories of Data and the Database," the authors explored the resurrection of interest in data histories, specifically tying this interest to the new ways in which data collection shapes modern life.[61] In a recent special issue of the history of science journal *Osiris*, the power of quantification was reassessed by many of the same authors in the current era of data mining, data hacking, and large-scale accumulation of data by private companies.[62] This collective effort to quantify, the authors argue, has even led to the creation of new statistical doppelgangers that move throughout the internet, "data doubles" or "algorithmic selves" that more and more stand in for our actual selves. In their malleability and transparency, these new "people" seem to resemble the "mere agents" necessary for AI success.[63] Rather than data collection "revealing" some essential "human level," these stories demonstrate that different forms of data collection have historically *produced* very different kinds of people, often reducing the concept of human thought and behavior to a form where it could be more easily mimicked by a machine.[64]

This recent work is important in updating the long history of quantification and data collection, and it is now beginning to find more relevance in

histories and ethnographies of AI.[65] For example, Stephanie Dick and Hunter Heyck have examined different elements of the life of Herbert Simon, the social scientist and AI visionary who combined a novel approach to data with the idea of thinking machines.[66] Two recent works by Rebecca Lemov and Jill Lepore have explored fantastical dreams from the 1950s to assemble complete knowledge in the social sciences, and Lemov's work in particular identified the link between data collection and modern selves.[67] Yet few of these transformative studies of AI go back before the twentieth century. Perhaps most surprisingly, even the flood of books on the "quantified self" movement barely mention data collection prior to the first electronic computers, focusing on modern anthropology and ethnography over historical links to the first quantifiers.[68] It is the argument of this book, however, that the reductive aspects of modern data collection in AI have far deeper roots and that the collective work of historians of data collection and the social sciences can provide the basis for an alternate history of AI. Rather than the "ivory tower" having "ignored" technology, the breadth of studies surveyed above provides a powerful challenge to the determinist narratives that have emerged from AI researchers and the first drafts of science journalism.

The Descent

When artificial intelligence is viewed through the lens of the history of science, the simple story of a "quest" begins to fade. As those familiar with the long history of social scientific theories of humanity are well aware, the history of the sciences of man has not been a steady climb to enlightenment but rather a long trail of disappointment and error, one that might better illuminate the future path of AI research than the triumphant histories of the past few decades. It is for this reason that the book is told as a story of *descent* rather than a quest. Like Charles Darwin's usage in his own fledgling science of man—*The Descent of Man* (1871)—the term is intended to be ambiguous, both denoting the literal description of how ideas from the social sciences were inherited by AI and also connoting a sense of perceived moral or social decline. As will be seen, the reductive aspects of the sciences of man and social sciences were repeatedly criticized by contemporaries. From the earliest religious critics of Descartes to the "interpretive revolution" of the 1970s, critics noted the dehumanizing or injurious consequences of reductive social science for the idea of intelligence. For some, "intelligence" itself managed to fall from the sky, as an idea once linked to divine providence became the possession of humans as a rote process of mechanical thought subject to imitation by machines.

Just as modern AI of the past fifty years can be seen either as a natural growth of ideas or a decline, the four hundred years from Descartes to modern machines can be viewed as either a story of progressive accumulation *or* degeneration. Indeed, even today researchers in AI warn that the most revolutionary approaches of the past few decades are approaching a moment of crisis and that the AI triumphalism of the late 2010s has waned, with the idea of building a mechanical mind replaced by the economic potential of generative AI. As Melanie Mitchell pointed out in a recent survey of the field, "the quest for robust and general intelligence . . . may be hitting a wall: the all-important 'barrier of meaning.'"[69] Although "meaning" would certainly be an important step for computer scientists in creating a true AI, it is also an elusive idea that has bedeviled *social scientists* for centuries.[70] Similarly, in a 2018 *New Yorker* article on the perils of "superintelligence," the MIT physicist Max Tegmark remarked on the challenges that await future work in the field, challenges that might seem quite familiar to those trained in the social sciences and humanities. Tegmark claimed, "to program a friendly AI, we need to capture the meaning of life."[71] If AI researchers who pursue this path need a guide to their future prospects, they would be well served to look into the long and troubled history of the sciences of man and social sciences to answer such questions.

This book will not quite attempt to capture the meaning of life, but it does cover a long history of efforts to reduce human thought and behavior to a state where they could be understood through science and mimicked by machines. In framing the history of the idea of artificial intelligence as the story of a descent, the book therefore traces what might be called a "genealogical" path rather than the "clear" and "tidy" history that is most common in internal accounts of AI.[72] Rather than a simple "origin" story for AI, it provides the history of a set of ideas and practices that have come to shape the context in which AI has been developed. Instead of a series of progressive discoveries, where each new thinker self-consciously builds upon the ideas of the past, the story told here is one of ironical and paradoxical fits and starts, where ignorance of the past, new discoveries in the natural sciences, and the stubborn barrier of human complexity combined to vex those looking to study human thought and behavior through the tools of the human sciences. In some cases, knowledge accumulates, but in the majority of the stories seen below, it collapses. What does appear to change, however, and what so concerned the critics of reductive social science, is that throughout this history, the broad possibility of human thought and behavior—and even the idea of intelligence

itself—became dramatically circumscribed. Ironically (again), while no one particular form of social science triumphed in the four hundred years covered by this book, many argued that actual human lives were transformed through the labor of scientific work and exclusive education in scientific methodologies. As in the case of artificial intelligence today, even when the sciences of man failed at the level of theory, critics maintained that in practice they nevertheless remained a powerful determinant of human lives.[73]

Reframing the story of modern AI—from a short history of a few technological triumphs to a long history of discarded social scientific theories—therefore requires a significant expansion in the sources traditionally used. This "deep history" of AI covers a wide range of reductive ideas that might be classified under the heading of "sciences of man." Among the creators of these sciences, the book includes discussion of René Descartes; the French Enlightenment thinkers Bernard de Fontenelle, Étienne Bonnot de Condillac, Baron d'Holbach, Claude-Adrien Helvétius, and the Marquis de Condorcet; the "father of economics," Adam Smith; nineteenth-century European social thinkers Adolphe Quetelet, William Stanley Jevons, Harriet Martineau, Charles Knight, Herbert Spencer, Alfred Russel Wallace, and Francis Galton; and twentieth-century American academic social scientists Frank Knight, Frank Hankins, William Ogburn, Milton Friedman, John Watson, Clark Hull, and Warren McCulloch. At the same time, the book includes a number of contemporaneous critics of these approaches, including French philosophes Jean le Rond d'Alembert, Denis Diderot, and Jean-Jacques Rousseau; the unclassifiable Julien Offray de La Mettrie; French reactionary monarchists Louis de Bonald and Joseph de Maistre; the British novelist Thomas Love Peacock; groundbreaking scientific thinkers Pierre-Simon Laplace and Charles Darwin; and twentieth-century academics Thorstein Veblen, Earle Eubank, Charles Ellwood, Wesley Mitchell, H.S. Jennings, Talcott Parsons, and John Tukey. While the dividing lines were not always so clear-cut between critics and proponents of a science of man—with many of the critics arguing for their own sciences of man—the overall discourse reveals that reductive social scientific theories held the potential in both theory and practice to drag human activity down to the simplistic degree of a machine.

Some of these names are well known to most readers (Descartes, Smith, Darwin), but some may be known only to historians of ideas and science (d'Alembert, Martineau, Jevons) or disciplinary specialists (Knight, Eubank, Tukey). Such a wide scope was both intentional and accidental. Many of the

sources were encountered in research for a book on the nineteenth-century statistician and social scientist Adolphe Quetelet—himself a key figure in the story—where it became apparent that the many debates that have arisen over AI in the past decade were foreshadowed in the contested histories of the sciences of man, particularly those sciences based in quantification, probability, and mathematical reduction. Rather than produce a study limited to the well-trodden paths of intellectual history, the book looks at lesser-known works and stories to reveal moments when the most reductive forms of social science emerged. In cases where the author is well known—like Descartes and Smith—the analysis goes beyond their most famous works to bring out ideas usually only discussed in specialist literature. For less well-known figures, it relies on letters, notes, and published work that has rarely received substantial discussion in *any* context outside of specialized monographs and academic journal articles. For the selection of critics, I have mostly looked at those figures who have emerged from *within* the discourse of the sciences of man, rather than the legions of artistic and literary challenges to these ideas that have stretched from the Romantics through Dada to the Situationists. While the selection was idiosyncratic and personal, the intent has been to locate and critically examine the practices and critiques of reductive social science that appear most relevant to the discourse of artificial intelligence over the past several decades.[74]

While the approach described above can certainly allow for insight into how human thought and action came to a point where it could be modeled by a machine, it does not allow for a full and exhaustive survey of the social sciences, or even capture every reductive approach to studying people in the past four hundred years. In order to keep the notes at a reasonable length, only secondary material most pertinent to the argument has been referenced, though many more works that influenced this book could have been included. As noted above, the book is predominantly concerned with ideas in Europe and America, but it contains relatively little from Germany, as science in the nineteenth-century German-speaking regions largely *challenged* reductive accounts of human action that relied on the tools of the natural sciences, following in part the warning of the philosopher Johan Georg Hamman (1730–1788) that "the sciences, if they were applied to human society, would lead to a kind of fearful bureaucratization."[75] Indeed, even by 1862 the physiologist and influential German methodologist Hermann von Helmholtz (1821–1894) could still note that in his country "the name of science was often denied" to the physical sciences.[76] And in the first few decades of the twentieth

century, the anti-reductive methodological practices of the "Germanic school" in economics and sociology was prioritized in American universities, whose first professors had often been trained in German seminars.[77] It was just these practices that needed to be overcome for professional social science to emerge in America. Similarly, anthropology—one of the primary social sciences of the twentieth century—is barely mentioned in the book, as the field spent much of the twentieth century *resisting* reductive accounts. While forerunners to anthropology like anthropometry, phrenology, and psychophysiology all contain startling reductions of human action and quantification, these stories have been well covered elsewhere and thus are only hinted at in what follows.[78] While the inability to include a broader range of stories is a liability for nearly any history with a long scope, the hope here is that those readers who know part of these stories are able to discover something new. Or, in the case of more recognizable names, they are able to see a new aspect to what seemed a familiar story.

It should also be noted that the approach described above is unlikely to reveal that modern AI researchers of the past seventy years self-consciously *based* their ideas on the sciences of man. In fact, the opposite may be true: modern AI successes and frustrations might be due to large swaths of ignorance in engineering and computer science departments about such histories. Historians and philosophers of science in particular have argued that the modern "scientific method" proscribes more historical and cultural approaches, instead offering what the historian of science Henry Cowles has recently called "an artificial, algorithmic scientific method."[79] For the philosopher of science Michael Strevens, science has in fact made great strides *because* it has been reduced to a simple set of narrow empirical theories, as he compares modern science to a process where a "simpleminded strategy" denies "students . . . the ability to think philosophically, theological, or aesthetically at all."[80] Although Cowles and Strevens do not mention AI and are largely focused on the natural sciences, their historical descriptions of the process of scientific methodology mirrors contemporary accounts of AI research. For example, in a recent critique, the entrepreneur and computer scientist Erik J. Larson noted that a "simple" approach to human intelligence is today a prerequisite for the field, claiming that "a dangerous simplification in the philosophical ideas about man and machine" is one of the chief obstacles to achieving a true AI.[81] Criticizing his fellow engineers for their limited approaches toward building thinking machines, Larson laments in *The Myth of Artificial Intelligence* that through "pulling down human intelligence, tying it to a definition more

amenable to computation, current thinking about AI jettisons a richer understanding of mind."[82] Although Larson does not look beyond the past half century to tell this story, his recent insider view of the field confirms the central thesis of this book of a long descent of reductionist and simplistic ideas in the social sciences as the key to the idea of thinking machines.

The process of "pulling down human intelligence" occurred in a wide variety of social scientific fields across several centuries. In order to provide some narrative coherence, the book has been organized chronologically and geographically, exploring simultaneously how both social thinkers and their critics articulated the process by which the sciences of man simplified its subject. Part 1, which largely covers France in the seventeenth and eighteenth centuries, begins with the influence of mechanical philosophy and an increasing interest in applying techniques in the mathematical and natural sciences to the study of human beings, an idea dubbed the "geometric spirit" by Fontenelle, the head of the French Académie des sciences. Such an idea inspired a host of Enlightenment sciences of man that assumed that social thinkers could provide the kind of order to human life that Isaac Newton had brought to the heavens. Chapter 1, "Intelligence Lost," explores the depths of Descartes's revolutionary mathematical and geometric approach to understanding human action, tracing its influence on French thinkers like Fontenelle and Condillac, as well as highlighting criticism from French philosophes like Rousseau, Diderot, and d'Alembert. Chapter 2, "At the Bleeding Edge," explores how the geometric spirit was applied in mechanical sciences of man developed by the French *salonistes* Helvétius and d'Holbach, as well as efforts by Diderot and d'Alembert to blunt the sharper edges of a mathematically influenced science of man. As seen by the end of the chapter, the certainties of a Cartesian science of man would be complicated by the probabilistic conclusions of Laplace, who imagined a "great intelligence" that could predict all future events based upon certain knowledge of the past. The final chapter of this section, "Warnings of a New Barbarism," recounts the severe criticism of the geometric spirit leveled by counterrevolutionary figures Bonald and Maistre. Although these thinkers had their own idea of a science of man, they railed against what they believed to be the denial of divine intelligence among the French Enlightenment *sciences humaines*, as well as the mundane mental labor it portended.

In the second part, the story expands to Europe in the nineteenth century, exploring how prominent sciences of man like social physics, social

mathematics, political economy, social Darwinism, and eugenics emerged in tandem with the growth of industry and machines in Europe to produce a new vision of the human subject, one which often ignored the Enlightenment focus on progress and equality that attended previous social thought. Chapter 4, "Progress to the Mean," explores how the idea of the "average man" was transformed in the statistical work of Condorcet, Quetelet, and Jevons. While Jevons combined the idea of a "social physics" with a mathematical approach to political economy, his work also revealed that the sciences of man did not merely *reveal* the workings of the human subject and mind but could *shape* the subject and mind as well. In chapter 5, "Tuning the Mind," Jevon's idea of an interventionist social science is explored in the work of earlier popular political economists Martineau and Knight, who sought to use the laws of the sciences of man as tools to instruct a population that did not always seem to follow the paths laid out by social physics or the "iron laws" of political economy. Through studying the early ideas of Smith and their deployment in Knight's and Martineau's popular pamphlets, the chapter explores how the interconnected ideas of political economy and real-life machines redirected attention to the *minds* of the workers at the machines. Chapter 6, "The Descent of Man (and Intelligence)," explores the fusion of social physics, political economy, and Darwin's *Origin of Species* in the emergence of social sciences developed by Spencer, Wallace, Galton, and Darwin himself, each of which would further challenge the Enlightenment focus on equality, progress, and human freedom that had motivated many of the earliest sciences of man. Here the geometric spirit was reborn as a form of biological reductionism, with even the idea of intelligence being reduced to a simple inheritable trait.

The final section of the book examines the development of formal academic social sciences in America in the twentieth century, many of which were strongly influenced by the evolutionary ideas of the previous century. While many eighteenth-century sciences had seen the abstract simplification of the human subject as the key to moral and political progress, and nineteenth-century social sciences suggested the need for political intervention and education in order to help people conform to these new laws, the move toward professional social science in fields like sociology, political economy, and psychology had at least three important consequences for the emergence of artificial intelligence in America. In the first place, it introduced a new kind of simplification and reduction: that of the social scientific researcher himself. Second, by projecting this newly professionalized ideal onto the human subject under examination, social scientists created forms of artificial

intelligence decades before the term itself was born. Last, by creating an idealized role for the researcher that stood outside of moral and political thinking, the model professional social scientist abandoned the early assumptions of the Enlightenment science of man that were explicitly guided by notions of equality, progress, and human freedom. At just the moment when the social sciences finally discarded their relationship to the broad methodologies and motivations of the Enlightenment, artificial intelligence was born.

In order to trace this transformation in the social sciences, chapter 7, "The Sacrifice and Rebirth of 'Man,'" examines the process of discipline formation in the world of sociology and economics in the first half of the twentieth century in America. It begins with the little-known story of Earle Eubank, a sociologist trained at the premier graduate school at the University of Chicago who, along with many others, struggled with accommodating his moral and political commitments to a new field of study. At the same time sociologists simplified the scope of their investigations, economics similarly became a "science" through the dismissal of methodologies that incorporated (explicit) moral thinking, a process that its critics worried reduced social scientific work to mental and spiritual drudgery. Chapter 8, "Social Science by Other Means," examines the consequences of such visions of social scientific work for the development of the idea of "intelligence" at the heart of AI. While two new prominent American social sciences of the 1930s and 1940s—behaviorism and functionalism—seemed to provide a new route to a science of man, their greater legacy would be in providing the central concepts of the human subject and environment for the pioneering AI work of thinkers like Turing and McCulloch. Not only did the earliest ideas of AI appear as the reflexive vision of the idealized professional (social) scientist, but such ideas also were able to thrive as the resources devoted to large projects of social science research were redirected toward computers in the competitive environment of the Cold War. Finally, chapter 9, "Second-Rate Mathematicians," examines a last attempt by the Princeton statistician John Tukey to guard against the increasing mathematical and quantitative reductionism in much social science research through the introduction of what he called *data analysis*. Tukey, one of the most neglected thinkers of the Cold War era, not only offered a vision of statistics and quantitative reasoning as autonomous subjects, warning against allowing data "to speak for itself," but he also offered an alternative to the simplified vision of the social scientist as quantifier. Years before the late philosopher of science Ian Hacking noted that ideas could "loop down" upon their creators, Tukey saw a similar process where mathematical and

quantitative simplification could have a recursive effect on social scientists and engineers.[83] If statisticians did not broaden their outlook, Tukey warned, quantitative thinkers themselves risked being reduced to a kind of artificial intelligence.

After examining how artificial intelligence emerged out of the same milieu as a host of failed and discarded social scientific theories, the book concludes by returning to the current fractured and uncertain world of AI, where the field seems closer to the final days of Cold War social science than it does to completing its quest for a truly intelligent machine. In its naivete, public excitement, financial backing, and philosophical positivism, AI has also ironically left the scientific study of human thought and action as fractured as the days when Alan Turing first questioned the ability to understand people through scientific induction.[84] As A.O. Scott, film critic for the *New York Times*, explained in a survey of AI films, even the hype surrounding innovations in generative AI in early 2023 revealed that AI at its best revealed only "the banality of sentience," as the modern version has failed to live up to the fevered dreams of futurists and science fiction writers.[85] Yet, as this book aims to show, the very real labor practices and biases inherent in centuries of social science may provide more relevant villains in modern AI than dystopian machines that attain some "human level." In our hopes and fears about artificial intelligence, we would do well to remember that the modern world of twenty-first-century AI—with its endless questions of meaning and purpose, hierarchy of human labor, and endemic bias—owes it most distinctive traits to its descent from the social sciences.