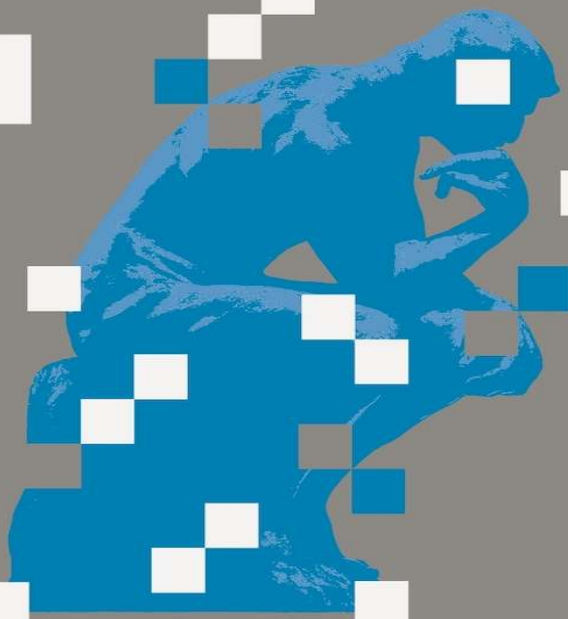
"Invaluable and necessary." —Ben Dickson, *TechTalks*

# Artificial Intelligence

## A Guide for Thinking Humans



Melanie Mitchell

PICADOR

# Artificial Intelligence

# A Guide for Thinking Humans

Melanie Mitchell

Picador

Farrar, Straus and Giroux

New York

**Thank you for buying this**
**Farrar, Straus and Giroux ebook.**

To receive special offers, bonus content,
and info on new releases and other great reads,
sign up for our newsletters.

Sign Up

Or visit us online at
us.macmillan.com/newslettersignup

For email updates on the author, click here.

To my parents,

who taught me how to be a thinking human, and so much more

# Prologue: Terrified

Computers seem to be getting smarter at an alarming rate, but one thing they still can't do is appreciate irony. That's what was on my mind a few years ago, when, on my way to a discussion about artificial intelligence (AI), I got lost in the capital of searching and finding—the Googleplex, Google's world headquarters in Mountain View, California. What's more, I was lost inside the Google Maps building. Irony squared.

The Maps building itself had been easy to find. A Google Street View car was parked by the front door, a hulking appendage crowned by a red-and-black soccer ball of a camera sticking up from its roof. However, once inside, with my prominent "Visitor" badge assigned by security, I wandered, embarrassed, among warrens of cubicles occupied by packs of Google workers, headphones over ears, intently typing on Apple desktops. After some (map-less) random search, I finally found the conference room assigned for the daylong meeting and joined the group gathered there.

The meeting, in May 2014, had been organized by Blaise Agüera y Arcas, a young computer scientist who had recently left a top position at Microsoft to help lead Google's machine intelligence effort. Google started out in 1998 with one "product": a website that used a novel, extraordinarily successful method for searching the web. Over the years, Google has evolved into the world's most important tech company and now offers a vast array of products and services, including Gmail, Google Docs, Google Translate, YouTube, Android, many more that you might use every day, and some that you've likely never heard of.

Google's founders, Larry Page and Sergey Brin, have long been motivated by the idea of creating artificial intelligence in computers, and this quest has become a major focus at Google. In the last decade, the company has hired a profusion of AI experts, most notably Ray Kurzweil, a well-known inventor and a controversial futurist who promotes the idea of an AI Singularity, a time in the near future when computers will become smarter than humans. Google hired Kurzweil to help realize this vision. In 2011, Google created an internal AI research group called Google Brain; since then, the company has also acquired an impressive array of AI start-up companies with equally optimistic names: Applied Semantics, DeepMind, and Vision Factory, among others.

In short, Google is no longer merely a web-search portal—not by a long shot. It is rapidly becoming an applied AI company. AI is the glue that unifies the diverse products, services, and blue-sky research efforts offered by Google and its parent company, Alphabet. The company's ultimate aspiration is reflected in the original mission statement of its DeepMind group: "Solve intelligence and use it to solve everything else."[1]

# AI and *GEB*

I was pretty excited to attend an AI meeting at Google. I had been working on various aspects of AI since graduate school in the 1980s and had been tremendously impressed by what Google had accomplished. I also thought I had some good ideas to contribute. But I have to admit that I was there only as a tagalong. The meeting was happening so that a group of select Google AI researchers could hear from and converse with Douglas Hofstadter, a legend in AI and the author of a famous book cryptically titled *Gödel, Escher, Bach: an Eternal Golden Braid*, or more succinctly, *GEB* (pronounced "gee-ee-bee"). If you're a computer scientist, or a computer enthusiast, it's likely you've heard of it, or read it, or tried to read it.

Written in the 1970s, *GEB* was an outpouring of Hofstadter's many intellectual passions—mathematics, art, music, language, humor, and wordplay, all brought together to address the deep questions of how intelligence, consciousness, and the sense of self-awareness that each human experiences so fundamentally can emerge from the non-intelligent, nonconscious substrate of biological cells. It's also about how intelligence and self-awareness might eventually be attained by computers. It's a unique book; I don't know of any other book remotely like it. It's not an easy read, and yet it became a bestseller and won both the Pulitzer Prize and the National Book Award. Without a doubt, *GEB* inspired more young people to pursue AI than any other book. I was one of those young people.

In the early 1980s, after graduating from college with a math degree, I was living in New York City, teaching math in a prep school, unhappy, and casting about for what I really wanted to do in life. I discovered *GEB* after reading a rave review in *Scientific American*. I went out and bought the book immediately. Over the next several weeks, I devoured it, becoming increasingly convinced that not only did I want to become an AI researcher

but I specifically wanted to work with Douglas Hofstadter. I had never before felt so strongly about a book, or a career choice.

At the time, Hofstadter was a professor in computer science at Indiana University, and my quixotic plan was to apply to the computer science PhD program there, arrive, and then persuade Hofstadter to accept me as a student. One minor problem was that I had never taken even one computer science course. I had grown up with computers; my father was a hardware engineer at a 1960s tech start-up company, and as a hobby he built a mainframe computer in our family's den. The refrigerator-sized Sigma 2 machine wore a magnetic button proclaiming "I pray in FORTRAN," and as a child I was half-convinced it did, quietly at night, while the rest of the family was asleep. Growing up in the 1960s and '70s, I learned a bit of each of the popular languages of the day: FORTRAN, then BASIC, then Pascal, but I knew next to nothing about proper programming techniques, not to mention anything else an incoming computer science graduate student needs to know.

To speed along my plan, I quit my teaching job at the end of the school year, moved to Boston, and started taking introductory computer science courses to prepare for my new career. A few months into my new life, I was on the campus of the Massachusetts Institute of Technology, waiting for a class to begin, and I caught sight of a poster advertising a lecture by Douglas Hofstadter, to take place in two days on that very campus. I did a double take; I couldn't believe my good fortune. I went to the lecture, and after a long wait for my turn in a crowd of admirers I managed to speak to Hofstadter. It turned out he was in the middle of a yearlong sabbatical at MIT, after which he was moving from Indiana to the University of Michigan in Ann Arbor.

To make a long story short, after some persistent pursuit on my part, I persuaded Hofstadter to take me on as a research assistant, first for a summer, and then for the next six years as a graduate student, after which I graduated with a doctorate in computer science from Michigan. Hofstadter and I have kept in close touch over the years and have had many discussions

about AI. He knew of my interest in Google's AI research and was nice enough to invite me to accompany him to the Google meeting.

# Chess and the First Seed of Doubt

The group in the hard-to-locate conference room consisted of about twenty Google engineers (plus Douglas Hofstadter and myself), all of whom were members of various Google AI teams. The meeting started with the usual going around the room and having people introduce themselves. Several noted that their own careers in AI had been spurred by reading *GEB* at a young age. They were all excited and curious to hear what the legendary Hofstadter would say about AI. Then Hofstadter got up to speak. "I have some remarks about AI research in general, and here at Google in particular." His voice became passionate. "I am terrified. Terrified."

Hofstadter went on.[2] He described how, when he first started working on AI in the 1970s, it was an exciting prospect but seemed so far from being realized that there was no "danger on the horizon, no sense of it actually *happening*." Creating machines with humanlike intelligence was a profound intellectual adventure, a long-term research project whose fruition, it had been said, lay at least "one hundred Nobel prizes away."[3] Hofstadter believed AI was possible in principle: "The 'enemy' were people like John Searle, Hubert Dreyfus, and other skeptics, who were saying it was impossible. They did not understand that a brain is a hunk of matter that obeys physical law and the computer can simulate anything … the level of neurons, neurotransmitters, et cetera. In theory, it can be done." Indeed, Hofstadter's ideas about simulating intelligence at various levels—from neurons to consciousness—were discussed at length in *GEB* and had been the focus of his own research for decades. But in practice, until recently, it seemed to Hofstadter that general "human-level" AI had no chance of occurring in his (or even his children's) lifetime, so he didn't worry much about it.

Near the end of *GEB*, Hofstadter had listed "Ten Questions and Speculations" about artificial intelligence. Here's one of them: "Will there be chess programs that can beat anyone?" Hofstadter's speculation was "no."

"There may be programs which can beat anyone at chess, but they will not be exclusively chess players. They will be programs of *general* intelligence."[4]

At the Google meeting in 2014, Hofstadter admitted that he had been "dead wrong." The rapid improvement in chess programs in the 1980s and '90s had sown the first seed of doubt in his appraisal of AI's short-term prospects. Although the AI pioneer Herbert Simon had predicted in 1957 that a chess program would be world champion "within 10 years," by the mid-1970s, when Hofstadter was writing *GEB*, the best computer chess programs played only at the level of a good (but not great) amateur. Hofstadter had befriended Eliot Hearst, a chess champion and psychology professor who had written extensively on how human chess experts differ from computer chess programs. Experiments showed that expert human players rely on quick recognition of patterns on the chessboard to decide on a move rather than the extensive brute-force look-ahead search that all chess programs use. During a game, the best human players can perceive a configuration of pieces as a particular "kind of position" that requires a certain "kind of strategy." That is, these players can quickly recognize particular configurations and strategies as instances of higher-level concepts. Hearst argued that without such a general ability to perceive patterns and recognize abstract concepts, chess programs would never reach the level of the best humans. Hofstadter was persuaded by Hearst's arguments.

However, in the 1980s and '90s, computer chess saw a big jump in improvement, mostly due to the steep increase in computer speed. The best programs still played in a very unhuman way: performing extensive look-ahead to decide on the next move. By the mid-1990s, IBM's Deep Blue machine, with specialized hardware for playing chess, had reached the Grandmaster level, and in 1997 the program defeated the reigning world chess champion, Garry Kasparov, in a six-game match. Chess mastery, once seen as a pinnacle of human intelligence, had succumbed to a brute-force approach.

# Music: The Bastion of Humanity

Although Deep Blue's win generated a lot of hand-wringing in the press about the rise of intelligent machines, "true" AI still seemed quite distant. Deep Blue could play chess, but it couldn't do anything else. Hofstadter had been wrong about chess, but he still stood by the other speculations in *GEB*, especially the one he had listed first:

> **QUESTION:** Will a computer ever write beautiful music?
>
> **SPECULATION:** Yes but not soon.

Hofstadter continued,

> Music is a language of emotions, and until programs have emotions as complex as ours, there is no way a program will write anything beautiful. There can be "forgeries"—shallow imitations of the syntax of earlier music—but despite what one might think at first, there is much more to musical expression than can be captured in syntactic rules.... To think ... that we might soon be able to command a preprogrammed mass-produced mail-order twenty-dollar desk-model "music box" to bring forth from its sterile circuitry pieces which Chopin or Bach might have written had they lived longer is a grotesque and shameful misestimation of the depth of the human spirit.[5]

Hofstadter described this speculation as "one of the most important parts of *GEB*—I would have staked my life on it."

In the mid-1990s, Hofstadter's confidence in his assessment of AI was again shaken, this time quite profoundly, when he encountered a program written by a musician, David Cope. The program was called Experiments in Musical Intelligence, or EMI (pronounced "Emmy"). Cope, a composer and music professor, had originally developed EMI to aid him in his own

composing process by automatically creating pieces in Cope's specific style. However, EMI became famous for creating pieces in the style of classical composers such as Bach and Chopin. EMI composes by following a large set of rules, developed by Cope, that are meant to capture a general syntax of composition. These rules are applied to copious examples from a particular composer's opus in order to produce a new piece "in the style" of that composer.

Back at our Google meeting, Hofstadter spoke with extraordinary emotion about his encounters with EMI:

> I sat down at my piano and I played one of EMI's mazurkas "in the style of Chopin." It didn't sound exactly like Chopin, but it sounded enough like Chopin, and like coherent music, that I just felt *deeply* troubled.
>
> Ever since I was a child, music has thrilled me and moved me to the very core. And every piece that I love feels like it's a direct message from the emotional heart of the human being who composed it. It feels like it is giving me access to their innermost soul. And it feels like there is *nothing* more human in the world than that expression of music. Nothing. The idea that pattern manipulation of the most superficial sort can yield things that sound as if they are coming from a human being's heart is very, very troubling. I was just completely thrown by this.

Hofstadter then recounted a lecture he gave at the prestigious Eastman School of Music, in Rochester, New York. After describing EMI, Hofstadter had asked the Eastman audience—including several music theory and composition faculty—to guess which of two pieces a pianist played for them was a (little-known) mazurka by Chopin and which had been composed by EMI. As one audience member described later, "The first mazurka had grace and charm, but not 'true-Chopin' degrees of invention and large-scale fluidity … The second was clearly the genuine Chopin, with a lyrical melody; large-scale, graceful chromatic modulations; and a natural, balanced form."[6] Many of the faculty agreed and, to Hofstadter's shock, voted EMI for the first

piece and “real-Chopin” for the second piece. The correct answers were the reverse.

In the Google conference room, Hofstadter paused, peering into our faces. No one said a word. At last he went on. “I was terrified by EMI. Terrified. I hated it, and was extremely threatened by it. It was threatening to destroy what I most cherished about humanity. I think EMI was the most quintessential example of the fears that I have about artificial intelligence.”

## Google and the Singularity

Hofstadter then spoke of his deep ambivalence about what Google itself was trying to accomplish in AI—self-driving cars, speech recognition, natural-language understanding, translation between languages, computer-generated art, music composition, and more. Hofstadter's worries were underlined by Google's embrace of Ray Kurzweil and his vision of the Singularity, in which AI, empowered by its ability to improve itself and learn on its own, will quickly reach, and then exceed, human-level intelligence. Google, it seemed, was doing everything it could to accelerate that vision. While Hofstadter strongly doubted the premise of the Singularity, he admitted that Kurzweil's predictions still disturbed him. "I was terrified by the scenarios. Very skeptical, but at the same time, I thought, maybe their timescale is off, but maybe they're right. We'll be completely caught off guard. We'll think nothing is happening and all of a sudden, before we know it, computers will be smarter than us."

If this actually happens, "we will be superseded. We will be relics. We will be left in the dust.

"Maybe this is going to happen, but I don't want it to happen *soon*. I don't want my children to be left in the dust."

Hofstadter ended his talk with a direct reference to the very Google engineers in that room, all listening intently: "I find it very scary, very troubling, very sad, and I find it terrible, horrifying, bizarre, baffling, bewildering, that people are rushing ahead blindly and deliriously in creating these things."

# Why Is Hofstadter Terrified?

I looked around the room. The audience appeared mystified, embarrassed even. To these Google AI researchers, none of this was the least bit terrifying. In fact, it was old news. When Deep Blue beat Kasparov, when EMI started composing Chopin-like mazurkas, and when Kurzweil wrote his first book on the Singularity, many of these engineers had been in high school, probably reading *GEB* and loving it, even though its AI prognostications were a bit out of date. The reason they were working at Google was precisely to make AI happen—not in a hundred years, but now, as soon as possible. They didn't understand what Hofstadter was so stressed out about.

People who work in AI are used to encountering the fears of people outside the field, who have presumably been influenced by the many science fiction movies depicting superintelligent machines that turn evil. AI researchers are also familiar with the worries that increasingly sophisticated AI will replace humans in some jobs, that AI applied to big data sets could subvert privacy and enable subtle discrimination, and that ill-understood AI systems allowed to make autonomous decisions have the potential to cause havoc.

Hofstadter's terror was in response to something entirely different. It was not about AI becoming too smart, too invasive, too malicious, or even too useful. Instead, he was terrified that intelligence, creativity, emotions, and maybe even consciousness itself would be too *easy* to produce—that what he valued most in humanity would end up being nothing more than a "bag of tricks," that a superficial set of brute-force algorithms could explain the human spirit.

As *GEB* made abundantly clear, Hofstadter firmly believes that the mind and all its characteristics emerge wholly from the physical substrate of the brain and the rest of the body, along with the body's interaction with the physical world. There is nothing immaterial or incorporeal lurking there.

The issue that worries him is really one of complexity. He fears that AI might show us that the human qualities we most value are disappointingly simple to mechanize. As Hofstadter explained to me after the meeting, here referring to Chopin, Bach, and other paragons of humanity, “If such minds of infinite subtlety and complexity and emotional depth could be trivialized by a small chip, it would destroy my sense of what humanity is about.”

# I Am Confused

Following Hofstadter's remarks, there was a short discussion, in which the nonplussed audience prodded Hofstadter to further explain his fears about AI and about Google in particular. But a communication barrier remained. The meeting continued, with project presentations, group discussion, coffee breaks, the usual—none of it really touching on Hofstadter's comments. Close to the end of the meeting, Hofstadter asked the participants for their thoughts about the near-term future of AI. Several of the Google researchers predicted that general human-level AI would likely emerge within the next thirty years, in large part due to Google's own advances on the brain-inspired method of "deep learning."

I left the meeting scratching my head in confusion. I knew that Hofstadter had been troubled by some of Kurzweil's Singularity writings, but I had never before appreciated the degree of his emotion and anxiety. I also had known that Google was pushing hard on AI research, but I was startled by the optimism several people there expressed about how soon AI would reach a general "human" level. My own view had been that AI had progressed a lot in some narrow areas but was still nowhere close to having the broad, general intelligence of humans, and it would not get there in a century, let alone thirty years. And I had thought that people who believed otherwise were vastly underestimating the complexity of human intelligence. I had read Kurzweil's books and had found them largely ridiculous. However, listening to all the comments at the meeting, from people I respected and admired, forced me to critically examine my own views. While assuming that these AI researchers underestimated humans, had I in turn underestimated the power and promise of current-day AI?

Over the months that followed, I started paying more attention to the discussion surrounding these questions. I started to notice the slew of articles, blog posts, and entire books by prominent people suddenly telling