Introduction

I suppose you could call me an AI critic, but it wasn't long ago that I was an absolute evangelist. I had my first real introduction to AI in 2015, when I transitioned from being an active-duty Army officer to a graduate student at Harvard. A friend shared a link to a blog that used colorful illustrations and language to describe how AI could lead to either a utopia or doom. Being a diehard optimist, I became instantly hooked on the AI idealist view. The doom side of things, most often addressed by "AI safety" or "AI alignment" researchers, seemed serious but manageable.

Like most introductions to the potential harms of AI, the blog focused on an AI system optimized to create … paperclips. This thought experiment, known as the paperclip maximizer, has become a cornerstone of discussions on AI safety. It goes something like this: Suppose you create a super-"intelligent" AI whose only job is making as many paperclips as possible for your paperclip-manufacturing facility. This seems harmless. But the AI would probably quickly realize that to achieve its mission, it'd need to ensure it couldn't be turned off. If it succeeded, we'd already be in some trouble, but the AI would eventually run out of metal to turn into paperclips, right?

Not so fast. Because the AI is extremely intelligent, it would explore other methods to produce paperclips. Metal, like everything, consists of atoms. So the super-intelligent AI could devise a way to break down different things into atoms, rearrange them into the types of atoms needed to form metal (e.g., add or remove protons), and then rearrange them into molecules that form metal, thereby obtaining plenty of metal for more paperclips. On a mission to optimize for paperclip creation over everything else, it would eventually break down most stuff (chairs, tables, concrete, glass, trees, etc.). This includes, tragically, humans. You. Me. Everyone. We'd just be atomic fodder for paperclips.

The scenario sounds absurd, but it illustrates a real problem: An AI system optimizing for a narrow goal without a broader understanding of human values could pursue it in catastrophically literal ways.

We'll revisit the alignment problem later.[1]
But on the other side of the issue: paradise!

[1] See Chapter 26 ("Artificial General Intelligence").

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001

AI could[2] potentially cure certain cancers, prevent and reverse global warming, end famines, develop ultra-efficient forms of energy production and storage, radically enhance education, turbocharge the economy, ensure the safety of all people, and, generally, allow humans to focus on whatever gives our lives meaning. Why worry about being turned into a paperclip when you can so easily envision a steady march toward utopia in our lifetimes?

I carried this excitement into graduate school, eagerly absorbing all proclamations of technological advancement from tech luminaries. Waymo's leader gave a TED Talk in 2015, in which he was hopeful they'd solve the self-driving vehicles problem within four years![i] In 2016, AI researcher Geoffrey Hinton told us that AI was so good at reading CT, MRI, and X-ray scans that we should stop training radiologists![ii] Also in 2016, the World Economic Forum estimated that AI would eliminate over five million jobs by 2020![iii]

Waymo's deadline came and went. Widely available, fully autonomous vehicles remain elusive a decade later.[3] Radiologists remain essential to medical diagnosis and treatment today. And while AI has indeed displaced some roles, the mass AI-induced unemployment predicted has yet to materialize.

Fully self-driving vehicles and AI radiologists would be tremendous technological achievements, but they would pale in comparison to achieving artificial general intelligence (AGI). Although there is no set definition for AGI, the term is usually associated with AI capable of performing all human cognitive tasks.[4] Unlike narrow AI systems designed for specific tasks, such as chess-playing programs or image recognition software, AGI would, in theory, match or exceed human intelligence across all domains.

My zeal for AI was renewed, perhaps even turbocharged, by my internships in the San Francisco Bay Area during law school. In 2017, I interned at LinkedIn in Silicon Valley. The internship felt like a dream. Everyone was kind and even generous. The weather was perfect, and somehow I landed an Airbnb in Los Altos Hills, an ideal area for someone who enjoys long walks on trails. Based on a recommendation from another LinkedIn intern, I began listening to venture capitalist (VC) podcasts while

[2] The word "could" here is doing a LOT of work. I'll revisit this in Chapter 27 ("Frameworks for Thinking about GenAI").

[3] Currently, Waymo only operates in eleven cities. It plans to expand to four more in 2026, making them tens of thousands of cities short of ubiquity. None of the cities are in areas with long and harsh winters or rainy seasons, and the vehicles almost exclusively drive on city roads rather than highways.

[4] It's unclear how much a distinction between cognitive and physical intelligence matters. If AI is as smart as humans, it should be able to design a robot that could perform any reasonable physical task, such as playing basketball, and it should be able to load itself into the robot to act as its "brain," providing the thinking necessary to guide the physical activity. When I talk about AGI, then, I mean AI that perform any task a human can, including physical tasks. Other definitions: passing the Turing Test (which requires a human to interactively distinguish whether text is produced by another human or by a machine), outperforming humans at "most economically valuable work," and AI being given $100,000 of capital and turning that into $1 million.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001

walking, such as those by Reid Hoffman and the VC firm a16z. These podcasts, hosted by the very people funding AI companies, naturally painted an optimistic picture of technological progress.

Marinating in these deeply self-interested views shaped my perception of technological progress and the astounding future ahead. Whenever I heard pushback from a tech critic, I would dismiss it reactively. If the critic didn't understand or believe how some technological prophecy would come to pass, it was because they lacked the insider knowledge of the CEOs and VCs. I deduced that the company leaders and their funders must know something that the critics didn't. Why else would they speak so confidently and optimistically?[5]

When I returned to law school that fall, I took a course called Law of Robots, which further stoked my interest. For my final paper, I wrote about how criminal liability might apply to AGI.[iv] The following summer, I interned in San Francisco, continuing my long walks and marathon podcast-listening sessions. During my final year of law school, I wrote another AGI-related paper, this time about whether an AI system could ever satisfy the constitutional requirements to become president. I concluded it could.[v]

Following graduation, I applied to OpenAI. They flew me to San Francisco for an in-person interview and espoused how mission-driven and research-focused they were, how tight-knit the company was (board game night every Thursday!), and how they planned to never go above about 100 employees.[6]

I ended up working in-house at an education technology company and began teaching myself to code via Codecademy. The learning pathway I chose? Machine learning, naturally. Feeling somewhat constrained in the role, which was limited to conducting contract reviews, I sought a new employer after a year and, figuratively, leaped into the world of autonomous vehicles. I bought into the idea that self-driving semi-trucks would be cruising across the nation within a couple of years, being more cost-effective and environmentally friendly than human-driven big rigs.

After pestering the chair of the Business, Government & Society Department at the University of Texas at Austin for three years, I secured a job teaching one course per semester there while working at an autonomous vehicle company. The department chair, with delightful foresight, asked if I'd like to teach a course on law, ethics, and AI. I happily and eagerly jumped at the chance.

The tech industry could not have asked for a more cheerful advocate, and I could not have asked for a more interesting starting time. My first law, ethics, and AI course launched less than two months after OpenAI introduced ChatGPT to the world in November 2022. While I had planned for the course to focus on machine learning, with each week covering a different topic – AI and medicine, AI and war, AI and finance, AI and the boardroom, AI and social media, AI and image recognition, and so on – I was too excited not to bring up fun examples ricocheting across the internet

[5] I'll revisit this in Chapter 2 ("Why GenAI Is Made").

[6] For reference, as of fall 2025, OpenAI is now estimated to have at least 4,000 employees.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001

from chatbots and image generators. It was all so new, neat, jaw-dropping, and fun! I specifically remember sharing with my class that DALL-E, OpenAI's image generator, had been trained on hundreds of millions of images. "Isn't that crazy?" I asked in awe. Think about it. *Hundreds of millions* of images processed to create a single AI system. Incredible.[7]

Initially, I did not pause to question the origin of the training data, how it was collected, or whether any of the sources had authorized its use to train OpenAI's models. When I lectured on the sources of data for training AI, I was primarily concerned with dataset representativeness to prevent or limit biased or inaccurate outputs. My focus was on AI that made loan-approval recommendations or on facial recognition that failed to identify darker skin tones. I don't think I mentioned copyright at all that semester.

My views on AI evolved over the course of the year. I left the autonomous vehicle company after alerting the federal government to what appeared to be the improper sharing of trade secrets with entities in China without proper agreements in place or the board's awareness. By a stroke of luck, I landed a job a few months later with the Allen Institute for Artificial Intelligence (Ai2).

Ai2's stated goal (at the time) was "AI for the common good." They had teams exploring the values embedded in AI and their implications for morality and ethics. They were drafting an AI ethics charter to express their commitment to the thoughtful and ethical development of AI. There was an AI & Society group focused on improving AI literacy, and the legal team was finalizing a new licensing agreement that aimed to strike a balance between openness to encourage research and acting responsibly. Best of all, the people I worked with all seemed genuinely caring, humble, intelligent, and kind. I'd found my home.

During my first month on the job, the general counsel (GC) assigned me a particularly interesting task: to help the team gain a deeper understanding of how copyright law intersects with AI. The GC had noted that the times were changing. It used to be that few people seemed to care if their data was scraped – that is, downloaded from a website, often without authorization – and used to train AI, possibly because they didn't know it was happening, but also perhaps because the scraping often benefited them. For instance, Google's web scraping improved search results, creating value for both users and website owners.[8] But Ai2, like most major AI organizations, had shifted its focus from narrow AI applications with clear use cases to generative AI (GenAI). At the time, I was all for it.

In May 2023, a month before I joined, Ai2 announced that it would create OLMo, "an open language model made by scientists, for scientists." The plan was to release a seventy-billion-parameter model in early 2024, trained on one of the world's greenest supercomputers. To put this in perspective, seventy billion parameters would

[7] I even offered extra credit for the student who created the most interesting image satisfying some basic criterion.

[8] See Chapter 20 ("The Economics of the Internet").

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001

make OLMo comparable in size to some of the most powerful language models available at the time. Ai2 stated that "the ethical and educational considerations around the creation of this model will also be openly shared and discussed to help guide the understanding and responsible development of language modeling technology." One of the lead researchers, Hannaneh Hajishirzi, was quoted in the press release saying, "OLMo will be the first language model specifically designed for scientific understanding and discovery."[vi] GeekWire accurately noted that "AI2 has also created an ethics review committee composed of both internal and external advisors to provide feedback throughout the process."[vii]

I was stoked. It was everything I wanted in an AI entity: cutting-edge research, transparency, and thoughtfulness about ethical considerations, as well as efforts to enhance society's understanding of AI and its implications through AI literacy. If there were good guys in AI, they were Ai2.

When it came time for me to research copyright law, a topic I had neither studied in law school nor worked on in my previous jobs, and to strengthen arguments that our use of copyrighted works was fair use, I was delighted to come across "Fair Learning" by Mark Lemley and Bryan Casey. The fair use doctrine, established in copyright law, allows for the limited use of copyrighted material without permission for purposes such as criticism, comment, parody, or research. However, its application to AI training data remains a hotly contested issue. Their paper made several arguments about why training machine learning systems on unauthorized content was likely fair use, and I ate it up, taking copious notes and briefing the legal team.[9] It was all coming together. We had ideals, we were pursuing them, and we had a legal argument to support them.

However, OLMo, as originally envisioned and announced, never materialized. The press release has since been removed from Ai2's website. Ai2 never published an ethics charter; the AI literacy program and the AI & Society group were deprioritized and eventually sputtered into irrelevance. Ai2 also switched from its proprietary responsible license to the permissive Apache 2.0 license and from the greenest supercomputer to GPUs in a decidedly un-green data center in hot, dry Austin, Texas. Ai2 also never publicly shared the ethical considerations behind building its LLMs, and the ethics review committee was quietly shuttered.

By that fall, I had begun asking why we were allowing the datasets and models to be used for commercial purposes. If our goal was to advance scientific research, why did anyone need to be able to make money off of it? To me, it was acceptable to use unauthorized works for scientific research, but to put in the legwork to collect and clean the data, train a high-performing model on it, and then allow anyone in the world to use any of it to generate revenue seemed … wrong. Under the text and data mining exemption of the Digital Millennium Copyright Act, universities were required to control access to copyrighted works, which made sense. I didn't

[9] I'll revisit copyright law in Chapter 11 ("An Introduction to Copyright Law").

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001

understand why research entities couldn't similarly control access to their AI artifacts when they used copyrighted works. In effect, such entities could be thought of as a free research and development department for companies.

Not long after, I made what I called my "Book of Questions." It was just a series of topics that I thought we hadn't put enough thought into. For example, I reviewed the results of several prominent surveys on AI, and they consistently show that Americans are more concerned than excited about AI development.[viii] I wanted to know whether and why Ai2 might develop technology that didn't align with the general public's views or desires, given our focus on "AI for the common good." More pointedly, I didn't understand why we didn't pursue a more democratic approach.

Given that all major GenAI companies were focused on creating AGI, some researchers began to wonder whether Ai2 also had that goal. In January 2024, someone asked the CEO about it at an all-hands meeting, and the CEO dismissed the idea, saying Ai2 couldn't possibly be working on AGI because he didn't even know how to define it.

At the break, in a room with hundreds of seats, the CEO happened to sit next to me. I wanted to revisit the question because public surveys had consistently shown that most people do not want AGI.[ix] "How do you define AI?" I asked.

"I don't know how to define it. It's amorphous," he replied.
"But Ai2 is working on AI?"
"Yes, of course."
"How can we work on AI without having a definition, but we can't work on AGI?"
He had no reply.

In late spring 2024, I began writing a law review article for the first time since law school. The idea was simple: review all the claims in GenAI litigation and tally them. I then provided a concise description of each claim and an example of how it could apply to GenAI.

I enjoyed writing the paper so much that I started another one shortly thereafter. It was about GenAI and freedom of speech under the First Amendment. A month later, I wrote two more. One was about GenAI and copyright law,[10] and the other was about GenAI and contract law.[11]

A few weeks after posting the papers on a preprint website, I was informed that the GC position at Ai2 was being eliminated and that I was being laid off because I had been there the shortest amount of time among all the lawyers. The legal team shrank from four to two members even as the risk of litigation rose, and lawsuits against GenAI companies increased. The reason – namely, to improve efficiency – was

[10] The paper is called David Atkinson, Unfair Learning: GenAI Exceptionalism and Copyright Law, 25 Chi.-Kent J. Intell. Prop. 43 (2026). Available at: https://scholarship.kentlaw.iit.edu/ckjip/vol25/iss1/5

[11] The paper is called David Atkinson, Putting GenAI on Notice: GenAI Exceptionalism and Contract Law, 120 Nw. U. L. Rev. Online 27 (2025), https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1357&context=nulr_online.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001

briefly provided at an all-hands meeting (neither the CEO nor the COO ever spoke with me directly about it). Apparently, either assessing the legal risks of novel technologies with novel applications was moving too slowly, or the legal team was so highly efficient, with too little work to keep us busy, that the CEO felt money spent on lawyers could be better allocated to marketing (a position posted shortly after we learned the legal team would be halved).[12]

The Book of Questions took on a life of its own in my academic work. I used ideas from it as a springboard for an open-source textbook that I work on with my students. We edited it each spring.[13] It was through developing that document that I began exploring the many, many ethical and legal issues that intersect with GenAI. Whenever I encountered a new area of concern, I added it to a list to explore in the book. Soon, the list of issues ballooned. A small sample includes:

- Hallucinations: when AI systems generate false or misleading information presented as fact
- Safety: the need to prevent AI systems from causing harm to users or society
- Disinformation/misinformation: the potential for AI to create and spread false information at scale
- Data scraping and consent: using copyrighted or personal data without permission
- Cybersecurity risks: AI systems as both tools for and targets of cyberattacks
- Environmental impact: the massive energy consumption required for AI training and operation
- GenAI oligopoly: the concentration of AI power among a few major companies
- Effect on the workforce: job displacement and transformation across industries
- Human labor for RLHF training: the often poorly compensated workers who train AI systems to be helpful and harmless.

Each of these issues represents a complex intersection of technology, law, and ethics that society is only beginning to grapple with.

By the time I left Ai2, I was firmly in the AI-critic camp. But this is not an anti-AI screed, and I am not anti-AI. I still believe AI, including GenAI, can be a force for tremendous good, but I'm far less inclined to think broad societal benefits are inherent or the most likely outcome. My previous rose-colored outlook seems more than naive to me today. It strikes me as childish.

GenAI – the technology behind popular chatbots like ChatGPT, Gemini, and Claude – is this decade's most consequential technological development. It has prompted multitrillion-dollar companies like Microsoft and Google to reorient their entire business strategies. At the same time, startups such as OpenAI, Anthropic, and

[12] If the second reason was the primary driver, an interesting fact is that, although Ai2 is at the forefront of AI research, the legal team did not utilize any AI tools. They were never necessary.

[13] I stopped this project once students started heavily relying on GenAI to do their writing, undermining the purpose of the exercise.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001

xAI (now a part of SpaceX) have seen their valuations surge into the hundreds of billions of dollars. Chip designers have also seen their valuations skyrocket in short order, with Nvidia, the company that provides the chips most widely used for AI, becoming the world's most valuable company.

Unfortunately, the legality of GenAI's rise is hardly on the general public's radar and is only slightly better known among legal scholars. I participated in the first seminar educating all Washington State district judges about GenAI, and none of them were familiar with how the technology worked or how it was developed. This is significant because if there is almost no GenAI literacy outside the tight-knit GenAI technologist community and legal specialists, society risks extending, expanding, and entrenching extraordinary legal privileges and exemptions in ways it has never done before.

The stakes could not be higher. We are essentially allowing a handful of companies to reshape fundamental aspects of how information is created, shared, and monetized, often using content created by others without permission or compensation.

This text aims to provide a friendly, engaging, frank, accessible, and informative introduction to GenAI and its foreseeable legal and ethical implications. I will rely on my law review articles, computer science papers, news articles, my open-source textbook, and insights from my work as in-house counsel at an AI research institute to inform the book's arguments. The chapters will be intentionally chunked into sections and kept relatively concise, making it easier for you to consume, understand, retain, and revisit the information.

I hope that by the end of the book, you will be equipped with the necessary knowledge and feel empowered to shape how GenAI affects you, rather than let GenAI companies choose how they will impact you.

The transformation is already underway. The question is not whether AI will reshape society, but whether we will have a say in how it reshapes society. This book is your invitation to join that conversation while there is still time to meaningfully influence its outcome.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 09:33:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009719605.001