

Review of ¹

If Anyone Builds It, Everyone Dies
Why Superhuman AI Would Kill Us All

Eliezer Yudkowsky and Nate Soares

Little, Brown and Company, 2025

272 pages, \$30 Hardcover, \$14.99 Ebook, \$24.99 Audiobook

Review by

Lance Fortnow



1 Summary

The lack of subtlety of the title reflects the book itself. Authors Eliezer Yudkowsky and Nate Soares lay out what they claim is an ironclad proof that if we develop Artificial Super Intelligence (ASI), then all humans will be wiped out. They then argue that we should immediately halt AI development, since we can't predict exactly when we might reach that deadly ASI threshold.

The authors are leaders of the Machine Intelligence Research Institute, co-founded by Yudkowsky in 2000 as the Singularity Institute for Artificial Intelligence. They define superintelligence as “a mind much more capable than any human at almost every sort of steering and prediction problem.”

The argument builds in chapters, and so let me express what the authors do in each.

Chapter 1: The authors argue that AI has gotten far closer to human reasoning faster than any of us had expected. Agreed.

Chapter 2: Modern artificial intelligence is “grown, not crafted.” It is trained on data and not by logic, and we don't understand the programs that get generated. Agreed.

Chapter 3: While the computers themselves don't have desires, they can simulate having desires, and we can reason about them as though they do. I struggle with this one and wrote a blog post² about it, which generated a lengthy discussion.

Chapter 4: AI will go beyond the desires we train it for and instead develop its own.

Chapter 5: AI's desires will not be compatible with human existence.

Chapter 6: A superintelligent AI is capable of killing us all.

That's Part I of the book. Part 2 describes an extinction scenario in which AI hides a killer virus in a new drug. Personally I would have gone with encouraging wars and revolutions and watching humanity kill itself off, but killer viruses also work.

Part 3 says we need to shut AI development down immediately and gives us directions for encouraging people to do so.

¹©2025 Lance Fortnow

²<https://blog.computationalcomplexity.org/2025/10/computers-dont-want.html>

2 A Call to Arms

The book provides mostly high-level arguments, often through stories. The authors give far more detailed reasoning on their website³, as they try to address every possible objection. On the other hand, despite this book being a bestseller and gathering many positive views, there hasn't been any significant push to slow AI development. If there are concerns about AI, they are mostly about how it affects our jobs, our intellectual property, and our education rather than as an existential threat. I would have liked to see the book balance the good that ASI could bring us with the risks, but if the authors truly believe the book's title, the good wouldn't do us any good if we weren't around to take advantage of it.

You should read this book for the reasons I did. As artificial intelligence's role in society continues to expand, it is essential for computer scientists, especially theorists, to understand its risks and consider various points of view, whether or not we agree with them.

While we have few tools to theoretically analyze the complexity of machine learning systems, we have developed a mathematical foundation of computing. We have an obligation to apply this knowledge to understand what computers can and cannot accomplish to help us understand the future of artificial intelligence and help shape it in the right direction. Our knowledge of computing gives us a unique frame to understand and interpret arguments about the future of artificial intelligence, such as the ones given in this book.

³<https://ifanyonebuildsit.com/resources>