

What is AI?

Zachary Peck
Assistant Professor of Philosophy
Ramapo College of New Jersey

What is AI?

Technology?

A science?

Or a kind of agent?



Brief history of AI

- 1956 – Dartmouth workshop, ‘artificial intelligence’ is coined
 - Many were confident that AI as intelligent as humans was just around the corner...
- 1980s – Logic-based AI: expert systems, knowledge-based systems, GOFAI
 - Many were confident that AI as intelligent as humans was just around the corner...
- Late 2000s – 2010s – Stats-based AI: machine learning, deep learning, neural networks
 - Many remain confident that AI as intelligent as humans was just around the corner...
- Late 2010s – 2020s – Generative AI: Language models, diffusion models, agentic AI
 - Many are now confident that AI is as intelligent as (some) humans and superintelligence is just around the corner...

Is AI going to kill us all?: A brief history

- I. J. Good (1960s) – coins the term “intelligence explosion” (aka, the singularity)
 - “the first ultraintelligent machine is the last invention that man need ever make...” (1965)
- Extropianism and transhumanism (1990s)
 - Increased discussion and focus on the possibility of creating machine superintelligence
- Founding of institutions devoted (in part) to the study of AI existential risk
 - Future of Humanity Institute (FHI) – 2005
 - Founding of Machine Intelligence Research Institute (MIRI) – 2005
- AI alignment (late 2010s)
 - A research paradigm aimed at ensuring potentially superintelligent AI systems are aligned with human values emerges
- Fragmentation of AI x-risk community (2020s)
 - Skepticism of Terminator-like scenarios
 - Increased attention given to social injustices entangled with AI industry
 - Increased focus on AI governance

AI as (an interdisciplinary) science

Initially, AI was conceptualized as an interdisciplinary academic project consisting of:

- Cybernetics, mathematics, computer science, philosophy, logic, psychology, cognitive science, linguistics, economics, etc.

At some universities, AI flourished as an interdisciplinary department or institute; while, at most universities, AI became a sub-discipline within computer science.

Today, there is a resurgence of interdisciplinary departments devoted specifically to the study of AI.

Philosophy of (interdisciplinary) science

Thomas Kuhn (1962) – *The Structure of Scientific Revolutions*

- Introduces the concepts of scientific revolutions, crises, and paradigm shifts

Research question – To what extent does contemporary AI (both machine learning and generative AI) constitute a scientific revolution, crisis, and / or paradigm shift?

Most discussion of within Kuhnian philosophy of science implies that scientific revolutions, crises, and paradigm shifts happen *within specific scientific disciplines...*

In my research, I aim to argue that we have misunderstood the significance of scientific crises. A scientific crisis challenges the very disciplinary boundaries according to which “normal science” operates.

AI as (a political) technology

Langdon Winner (1980) – “Do artifacts have politics?”

- Yes, technologies are inherently political.
- Example: Robert Moses and the low-hanging overpasses in Long Island

LLMs instantiated as chatbots is an entirely contingent choice based upon a variety of a social, political, and economic choices...

- “Move fast and break things” (Zuckerberg) – *Move Slow and Upgrade* (Selinger and Cahn)
- How to monetize a technology? – The social media model

Ethics of (a social, political, environmental, and economic) technology

The ethics of AI cannot adequately operate as a niche topic in applied ethics.

Rather, AI has direct social, political, environmental, and economic consequences. Thus, the ethics of AI is best understood as a microcosm of all applied ethics and political philosophy...

- Crawford (2021), *Atlas of AI* – AI is a planetary technology with profound social, political, and environmental consequences.
- Noble (2018), *Algorithms of Oppression* – AI reinforces and exacerbates systemic patterns of injustice
- Zuboff (2019) *The Age of Surveillance Capitalism* – AI as an industry is driven by the need to extract and monetize data
- By far, not a comprehensive list...

Research question – What does this kind of normative entanglement tell us about how we ought to think about applied ethical and political philosophical problems?

AI as (a group) agent

With the rise of agentic AI, there is an increasing attention given to whether or not we should think of AI systems as political, legal, or moral agents.

List (2021) – “Group agency and AI”

- Groups may be treated as agents due to a variety of cognitive, moral, and legal considerations.
- AI systems are similar to groups across many of these considerations.
- Therefore, we have good reason to treat some AI systems as (at least potential) moral and legal agents.

List (2021) on Group Agency and AI

“In fact, group agents can be viewed as special cases of AI systems, where the “hardware” supporting their artificial intelligence is social rather than electronic... As functionalist philosophers of mind have long argued, agency is a multiply realizable phenomenon: it can be realized by different “hardware systems”. The foregoing discussion illustrates that agency admits at least three different kinds of hardware: biological, as in the case of human beings and nonhuman animals; electronic, as in the case of robots and other AI systems; and social, as in the case of group agents. The boundaries between these three kinds of hardware may become increasingly blurred with the arrival of new (often controversial) technologies, from biological robotics and human enhancement to technologically augmented social systems.” (9)

Cyborgian group agents

Put simply, I think List focuses on the wrong question...

What's more important than whether or not groups or AI systems can be treated as agents in isolation from one another is how the creation and integration of AI systems into already existing groups and social networks changes that structure of those groups and networks.

I argue that:

1. AI systems are both a manifestation of, and a kind of, group agency.
2. Cyborgian group agents pose novel social, moral, and political challenges.

Is AI going to kill us all?

I argue that this question ought to be replaced with the following:

To what extent do emerging cyborgian group agents pose an existential risk to humanity?