The Ethics and Effective Use of AI

Scott Samuelson
Extension and Outreach
Philosophy and Religious Studies
Iowa State University

2024 ISAC Annual Conference
Wednesday, August 21, 2024: 2:30 pm – 4:30 pm

AI – Avian Influenza

So glad you're interested in Avian Influenza and community development . . .*

*This presentation is certified organic.

The blind are using autonomous cars driven by AI to get around on their own.

See: https://www.youtube.com/watch?v=X_d3MCkIvg8

Joshua Brown was killed when his self-driving Tesla crashed into a truck that the AI could not distinguish from the sky.

AI bomb-disposal robots can defuse or detonate deadly devices without risking human lives.

In South Africa, an anti-aircraft weapon lost control when directed by AI and sprayed cannon shells, killing 9 and wounding 14.

Thought River uses an AI program that reads legal contracts, answers key questions about them, and suggests next steps for a small fee.

See https://www.thoughtriver.com

Eric Loomis was sentenced to six years in prison because an AI-driven risk assessment identified him as a "high risk to the community."

See Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016).

IBM’s Watson diagnosed a patient with a rare leukemia that human doctors couldn’t pinpoint initially.

See www.asianscientist.com/2016/08/topnews/ibm-watson-rare-leukemia-university-tokyo-artificial-intelligence/

Data used to train an AI to predict the need for high-risk care led the AI to prioritize white patients over black patients with the same level of illness.

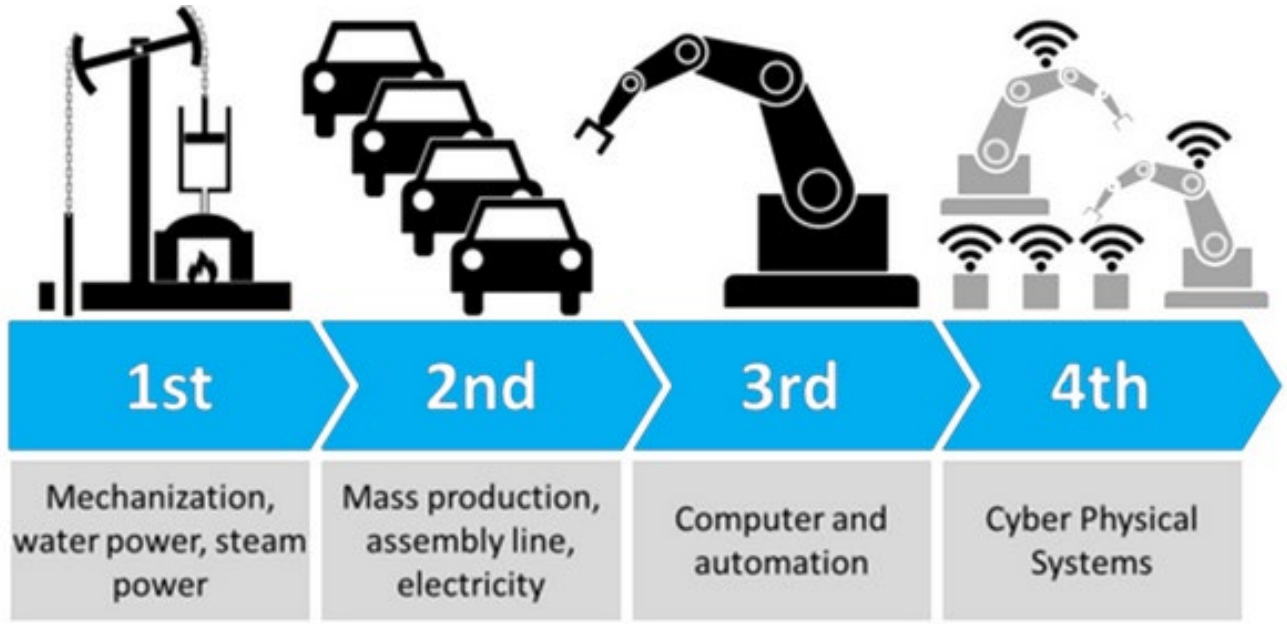
Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366, no. 6464 (2019): 447-453.

The Translational AI Center at ISU is creating digital copies of plants in order to run hundreds of years of experiments with multiple variables to help farmers deal with climate change.

In one class of philosophy students, at least a third were using ChatGPT to write their assignments for them.

This massive (ongoing) technological revolution will be an **ecological** shift.

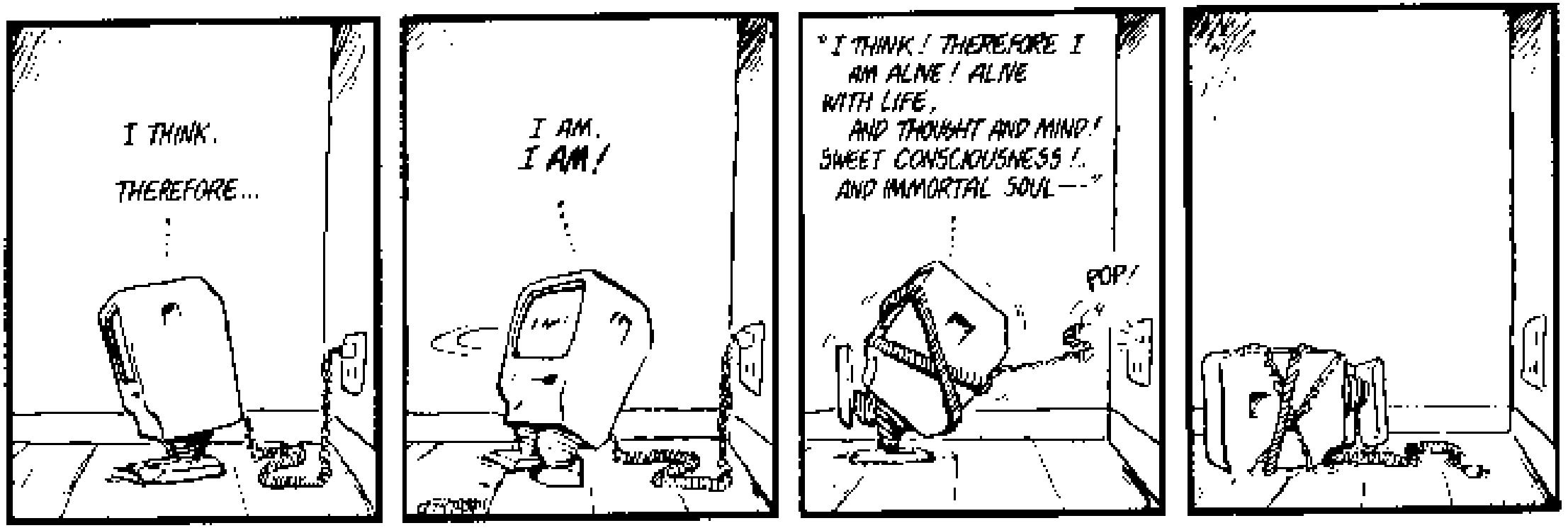
How can we be thoughtful in navigating the changes?



AI and Philosophy

Today let's think big so we can think better.

- I'm not here to hand out answers: I'm here to try to find the right questions!
- I'm not pro-AI or anti-AI.
- Luckily, we can draw on 2,400 years of thinking about AI!
- We're living through a thought-experiment!



A Deep Blue Parable

1985: Gary Kasparov simultaneously plays the best 32 chess computers in the world. His record is 32-0.

1996: Kasparov defeats Deep Blue, an IBM-powered chess computer.

1997: Deep Blue defeats Kasparov. “The Brain’s Last Stand” —*Newsweek*

2005: Playchess.com hosts an utterly open tournament.

The winning team is a pair of amateur American chess players using three computers at the same time.

Kasparov: “Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process.”

https://www.nybooks.com/articles/2010/02/11/the-chess-master-and-the-computer/

What’s the state of chess now?

Harry Goldstein: “Nowadays, chess newbies can log in and play engines that far exceed their own abilities, learning strategies and moves in days or weeks that in the past might have taken months or years. Engines can also help neophytes and grandmasters alike analyze their own games to give them an edge against human opponents.”

https://spectrum.ieee.org/ai-tools

Three Roads

Substitution: What things might AI do instead of us?

Collaboration: What things might be enhanced by human-AI collaboration?

Certified Organic: What things is it crucial for humans to do without AI? *Who are we anyway?*

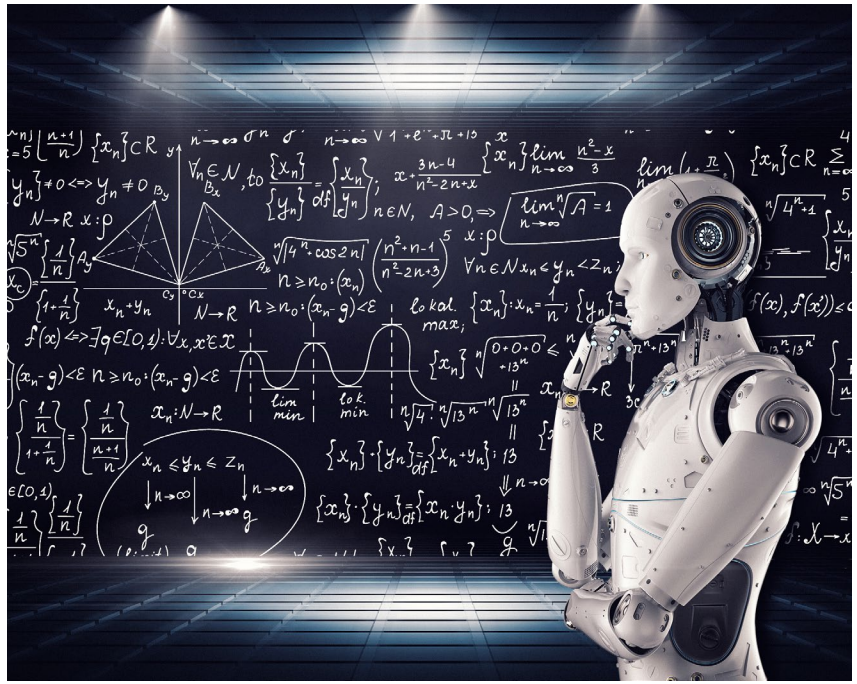
Part One

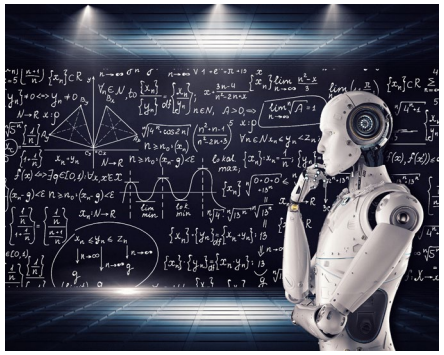
What is AI?
(Actually, nobody really knows!)

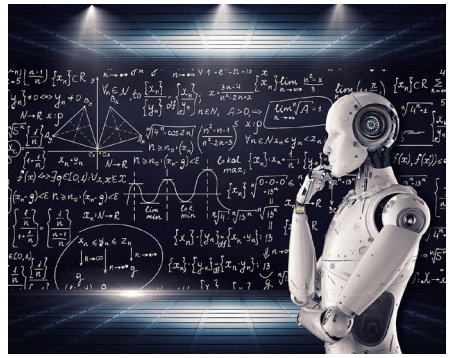
What is AI?

U.S. National Artificial Intelligence Initiative Act of 2020

A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments ***with sufficient reliability***.







Narrow AI (Weak AI?): An AI system designed to do a specific task or set of tasks – e.g., play chess, perform language-based tasks, drive a car.

General AI or AGI (Artificial General Intelligence/ Strong AI), a.k.a. Human-Level AI or Superintelligence: AI that could perform a number of tasks, including ones for which it was not designed – e.g., drive a car and answer questions and play chess. It's an ongoing debate the extent to which AGI has been achieved.

GOFAI

- Good Old-Fashioned AI: classic symbolic AI, where a specific task for a specific environment has been programmed into a machine – e.g., floor-cleaning robots. The problem with GOFAI is that it's hard to program in everything you might encounter in most situations or tasks.

Machine-Learning: a goal is programmed into the machine, and it can figure out how to achieve it.

Supervised Learning: AI is given specific feedback to learn to do its task, like teaching through flashcards. (Input and output variables are given. It learns from labeled data.)

Unsupervised Learning: AI finds previously unspecified patterns and works from those patterns, like figuring out things on your own. (Only input data are given. It learns from unlabeled data.)

Human vs. Machine

- AI outperforms us on any number of tasks already.
- What are some things it can’t do or doesn’t do well?

AI lacks common sense.

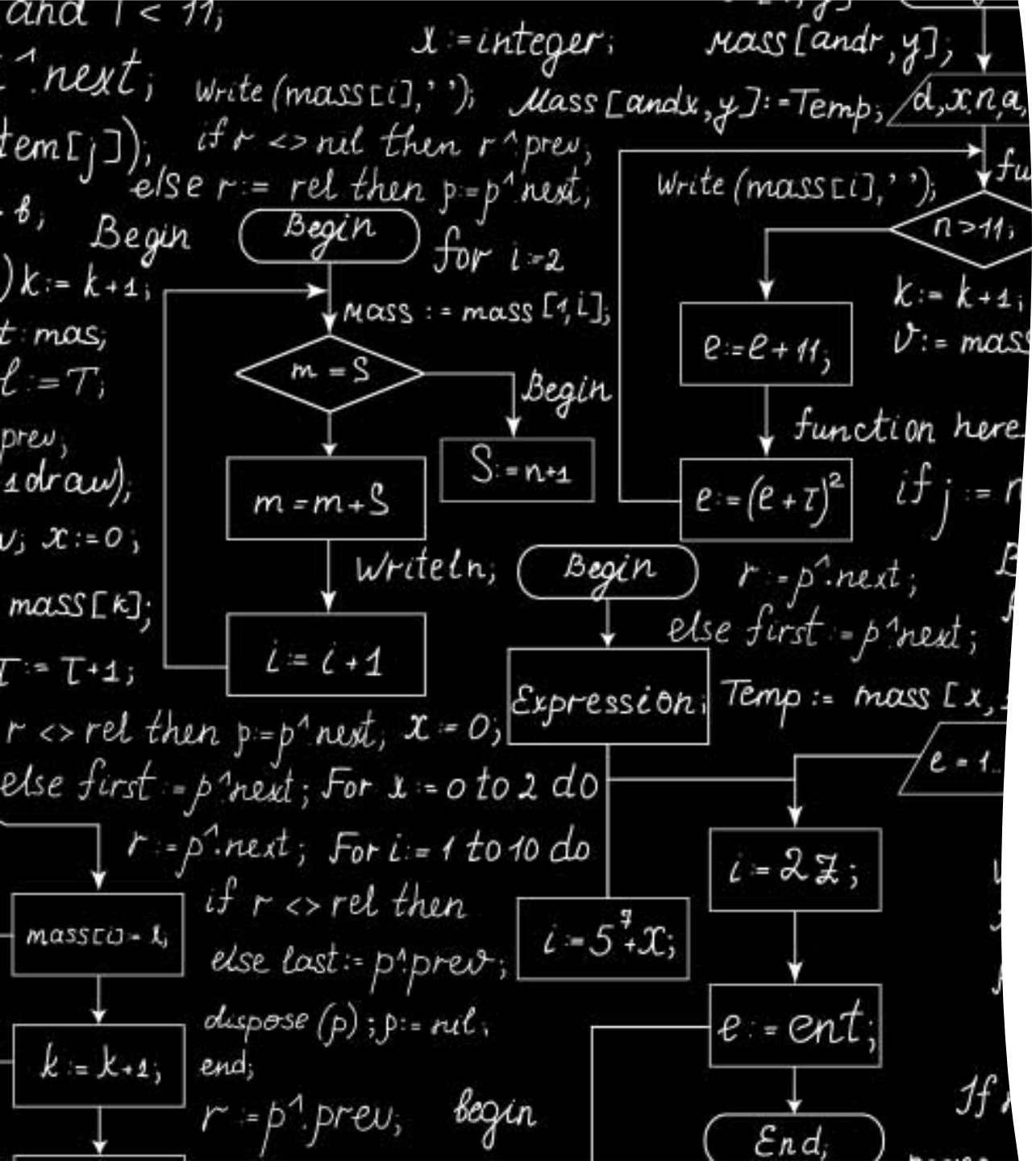
AI makes things up (“hallucinates”).

AI’s creativity isn’t exciting or revolutionary.
AI (moves quickly but) isn’t a quick study.

AI at best can only fake EI (Emotional Intelligence).

How is AI made?

Very roughly speaking . . .



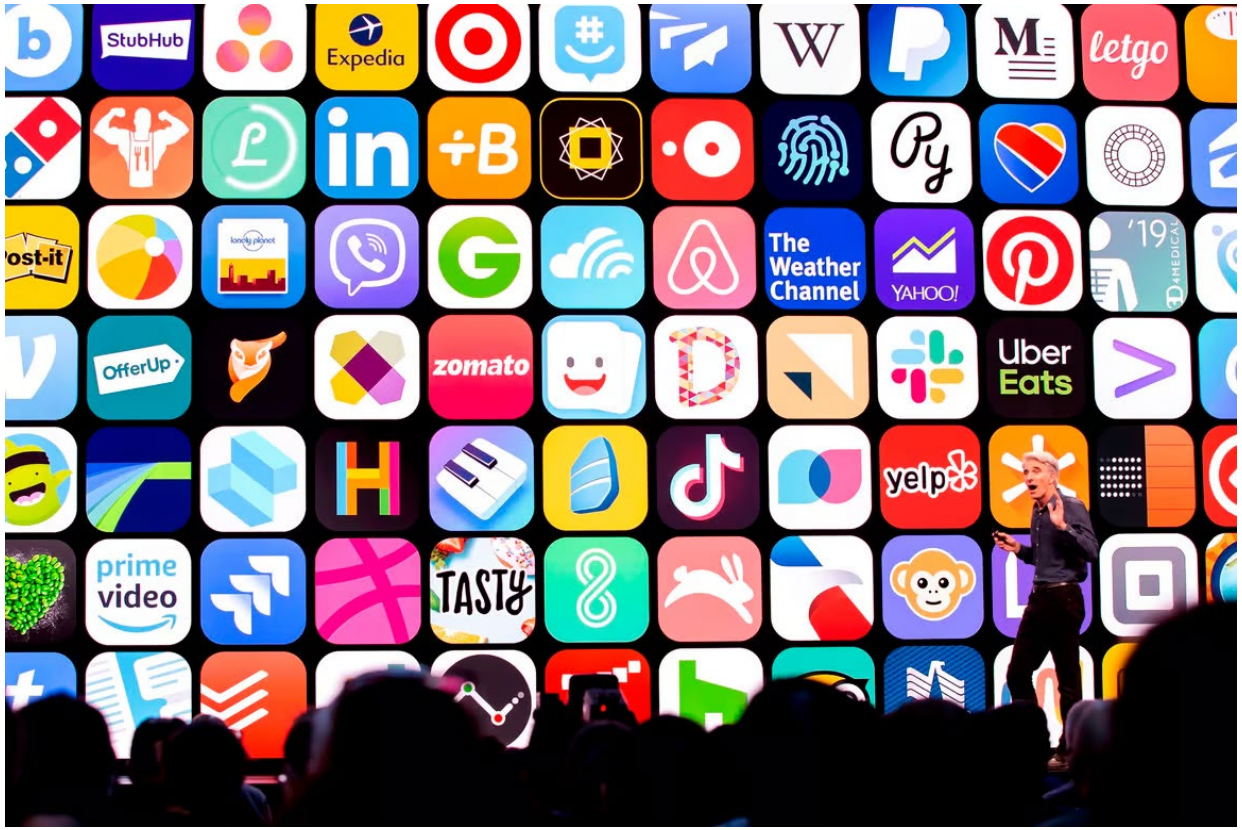
Development

Algorithms are developed that will apply to data and generate AI's intended behavior. This development is done in a research environment (academic, corporate, or government) by people with doctorates in mathematical fields.

Data Training

AI models are trained by data scientists, engineers, computer scientists, analysts, or community members who may or may not have advanced mathematical expertise. This training often requires considerable energy use.

Production/ User Interface



We get AI products, usually by powerful sponsoring organizations with a profit motive.

Open-source AI?

Brian Eno: “Artificial Intelligence is something that I've been interested in for quite a long time, I have several friends who are working in that area. And I'm not frightened of it, I’m frightened of the people who currently control it. Like the NSA and so on.”

“Open-source AI development offers three primary benefits: transparency, community-driven ethics and a counter to monopolistic control.”

https://www.rollingstone.com/culture-council/articles/beyond-corporate-ai-why-we-need-open-source-revolution-1234867419/

Part Two

Will AI take our jobs, ruin civilization, and then kill us all of?

(Or how might we not be used by it?)

The King Midas Problem

The problem of AI successfully pursuing a goal at the expense of a host of other goods – potentially to disastrous ends. This is especially a worry about AGI. But it can also do damage with narrow AI.

Over-trust in AI

We can put too much trust in AI (even though it's not particularly intelligent!)

Deskilling

As we become reliant on AI, we lose the skills to be good stewards of it and/or lose the skills that give us a way of making a meaningful contribution.

Mass Unemployment

The fear is that AI will render far more jobs irrelevant than create new ones.

Empowerment of Evil

AI systems can be used or hacked by bad actors to wreak havoc . . . or big tech itself is sometimes a bad actor . . .

Case Study: Transportation

Two Boeing 737 MAX airplane crashes (Lion Air from Jakarta to Indonesia and Ethiopian Airlines from Ethiopia to Kenya) in 2018–2019 occurred because of AI sensor malfunctions (**problem 1: AI makes mistakes**) and the pilots weren't able and/or weren't trained to override them (**problem 2: humans make mistakes**).

What will happen as these AI systems get better?

Risk 1: Pilots/drivers are deskilled.

Risk 2: Pilots/drivers become less vigilant.

Risk 3: Far fewer pilots/drivers are needed.

Risk 4: Bad actors can hack the system to wreak havoc.

Risk 5: People can figure out ways of fooling AI (e.g., a security team tricked a Tesla self-driving car into switching lanes by placing stickers on the road).

https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf

Another risk!

We don’t use AI even when it could be far safer. Needless lives are lost.

Does AI give us an opportunity to rethink who we are?

"Every assistant is a tool taking the place of several tools—for if every tool were able to perform its particular function when it was given the order or realized that something had to be done (as in the story of Daedalus' statues or Hephaestus' tripods which Homer describes as 'entering the assembly of the gods of their own accord'), so that shuttles would weave cloth or harps play music automatically, then master craftsmen wouldn't need assistants, nor masters slaves."

Aristotle, Politics

Can AI enhance our freedom?

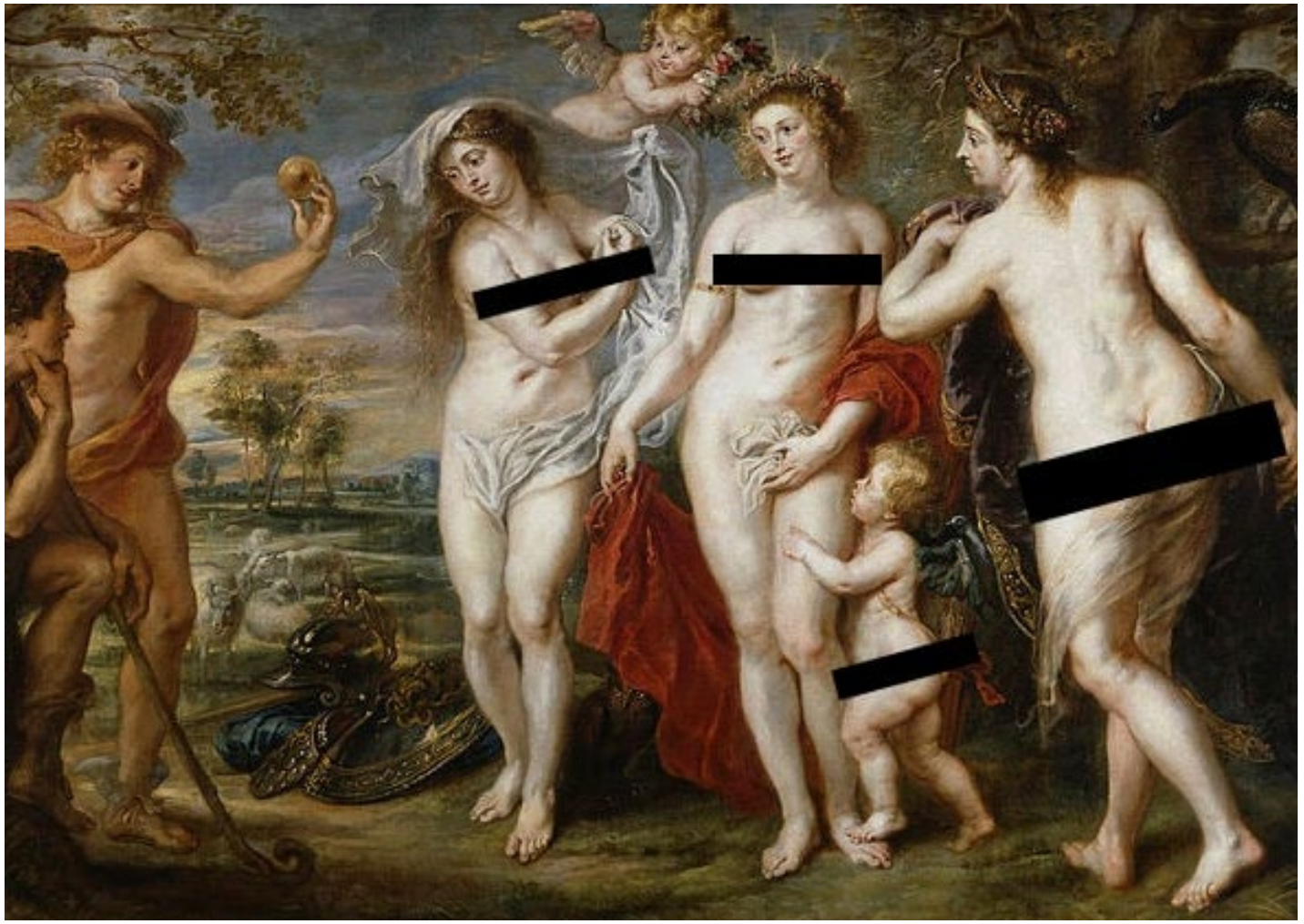
AI can do certain tasks for you. When should it? When is a task choice-worthy?

What unique contribution can we make?

How can we rethink what we do to offload certain tasks onto AI and enhance the human things?

Part Three

What ethical concerns should be uppermost in our mind as we move forward with AI?



Privacy

App designer: “I realized that x-ray glasses are possible!”

The result: DeepNude, where uploaded pictured of clothed women can be transformed into nude pictures of those same women.

https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman

How can AI violate our privacy?

AI can uncover or predict private information (e.g., through facial recognition) without our consent.

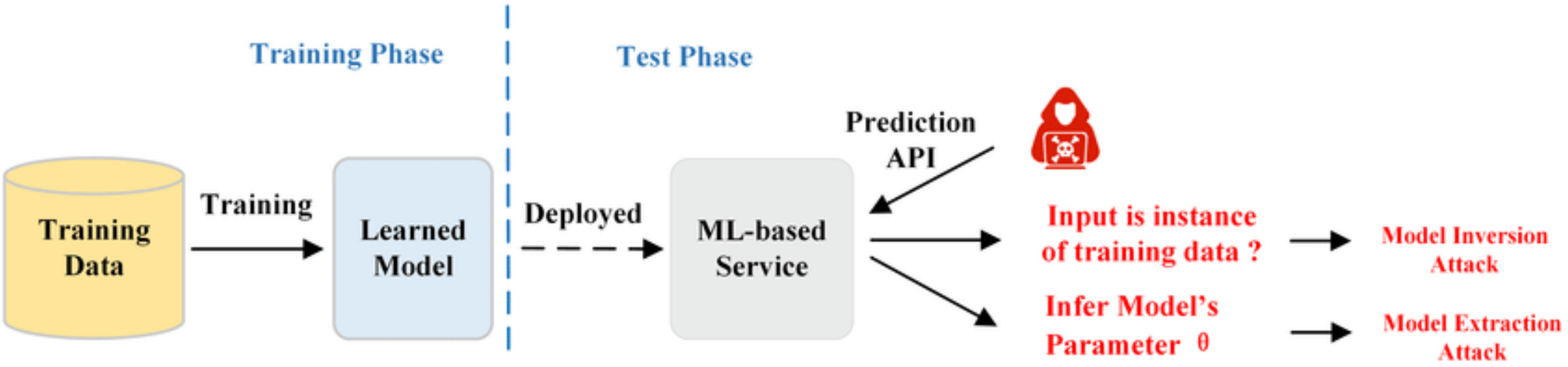
https://thelensnola.org/2018/10/24/months-after-end-of-predictive-policing-contract-cantrell-administration-works-on-new-tool-to-id-high-risk-residents/

How can AI violate our privacy?

AI can make phishing scams far more effective.

Model inversion attacks

Model inversion is a machine learning security threat that involves using the output of a model to infer some of its parameters or architecture.



Data Hoarding/ Data Use

AI incentivizes data collection and hoarding.

The Economist declared, “The world’s most valuable resource is no longer oil, but data.”

https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Amendment III

No Soldier shall, in time of peace be quartered in any house, without the consent of the Owner, nor in time of war, but in a manner to be prescribed by law.

Amendment IV

The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

Back to the future?

Bias

A-level AI?

Because of the pandemic in 2020 the UK used an algorithm to give grades based on teachers' assessments, individuals' performance on practice exams, and schools' exam track record in previous years. The result: many underprivileged students received lower scores than their teachers expected. Acceptances were limited or sometimes revoked due to the scores.

The big problem: **Technology can seem like it will fix social/political problems, even though such problems require social/political solutions.**
The specific problem: **AI can seem like it's unbiased/objective when it's not.**

In fact, AI can scale up whatever problems we already have. Bias in, bias out.

In predicting, AI relies on past information, which can further disadvantage the disadvantaged.

Self-Driving Responsibility

In 2019, during a test run of a self-driving Volvo for Uber, a woman who was wheeling her bike across a road in Tempe, Arizona was hit by the car. The test driver was streaming an episode of The Voice (unclear whether watching it or just listening).

Who's responsible?

The test-driver?
The pedestrian?
The AI designers?
Uber?
The government?
The self-driving car?

SPECIMENS OF THE ENGLISH SCHOOL: NO. V. PAINTED BY S. A. HART, R.A.

OTHELLO AND IAGO.

16

Truth, Transparency, Trust

The philosopher Daniel Dennett:

"Through AI we will have created the viruses—the mind viruses, the large-scale memes—that will destroy civilization by destroying trust and by destroying testimony and evidence. We won't know what to trust."

Education

Those with knowledge will likely find AI tools useful in their hunt for knowledge. But how will those with undeveloped skills and knowledge be able to assess what's generated by AI?

Resource Depletion

One estimate: ChatGPT gulps up 500 milliliters of water (close to what's in a 16-ounce water bottle) every time you ask it a series of between 5 to 50 prompts or questions.

Does the benefit of AI outweigh the amount of resources it consumes to run it?

Part Four

Can AI aid in community development?

AI and Community Development

- Decision Making
- Public Services
- Healthcare Access
- Education
- Community Engagement
- Economic Development
- Sustainability
- Disaster Preparedness and Response

Decision Making

Needs Assessment: AI can analyze large datasets to identify the needs of a community. *BUT how can we do this ethically and fairly and smartly; and how can we protect people's data from wrongful use?*

Predictive Analytics: AI can predict future community challenges based on historical data, allowing for proactive planning and resource allocation. *BUT how can we ensure that we're not reproducing old problems?*

Public Services

Smart Infrastructure: AI can optimize the management of utilities like water, electricity, and waste. Smart sensors and AI-driven systems can reduce waste, improve efficiency, improve safety, and lower costs. BUT how will we allocate responsibility with these optimizations?

Healthcare Access: AI can help in diagnosing diseases, managing patient data, and optimizing healthcare services, especially in underserved areas. BUT how can we keep what's best about human contact in health care and ensure that the care is high quality?

Education: How do we balance using AI, learning how to use AI, and learning the knowledge and skills that enhance our humanity and make us creative and responsible?

Community Engagement

“Temperature” Analysis: AI can analyze social media and other platforms to gauge public sentiment, helping leaders understand community concerns and respond effectively. BUT how can we dig deeper than what’s just online?

Chatbots and Virtual Assistants: AI can provide residents with 24/7 access to information about community services, events, and resources, improving accessibility and responsiveness. BUT how do we take responsibility and maintain the human touch?

Economic Development

Business Support: AI can help businesses with market analysis, customer engagement, and financial management. BUT how can we ensure that small business get the same benefits?

Sustainable Development

Environmental Monitoring: AI can monitor environmental factors like air quality, water levels, and wildlife populations. BUT how can ensure community input into major decisions?

Energy Efficiency: AI can optimize energy use in homes and public buildings, reducing costs and environmental impact.

Disaster Preparedness and Response

Early Warning Systems: AI can improve the accuracy and timeliness of disaster predictions, such as floods or tornadoes, allowing communities to prepare and respond more effectively. BUT how can we ensure that we have the social networks for real responses to disasters?

Resource Allocation: During disasters, AI can optimize the distribution of resources like food, water, and medical supplies. BUT how can we maintain the important human connections in our response?

Community Vitality

At its best, AI can help with various goals we have. But even at its best, AI won't make life worthwhile. How can we be the kind of communities that people flourish in? What are the things that AI will never do that make a good life and a flourishing state?

Some of my hopes for this presentation . . .

It's helped you to think more precisely about AI.
It's added to your vocabulary for making decisions about AI.
It's encouraged you to see AI as a human thing, full of promise and peril like everything else we deal with.
It's given you the groundwork for using and NOT USING AI thoughtfully, creatively, and ethically.
And maybe it's reminded you of the importance of human thinking.