What’s up with AI

Talk for the Topos Institute

October 30, 2025

Terry Winograd

Professor Emeritus of Computer Science

Stanford University

What are my concerns with AI today?

- It’s easy to get lost in the hype about AI as savior and AI as apocalypse – the boosters vs. the doomers
- This speculative long-term focus diverts our attention from the very real problems that we are already facing and will in the near future.
- I will focus on some of those problems
  - Income disparity and the economics of world populations
  - Erosion of trust in institutions, media, and information
  - Displacing human interactions and relationships
  - Surveillance
  - Depersonalization of violence
  - Abandoning responsibility
  - Fostering nihilism

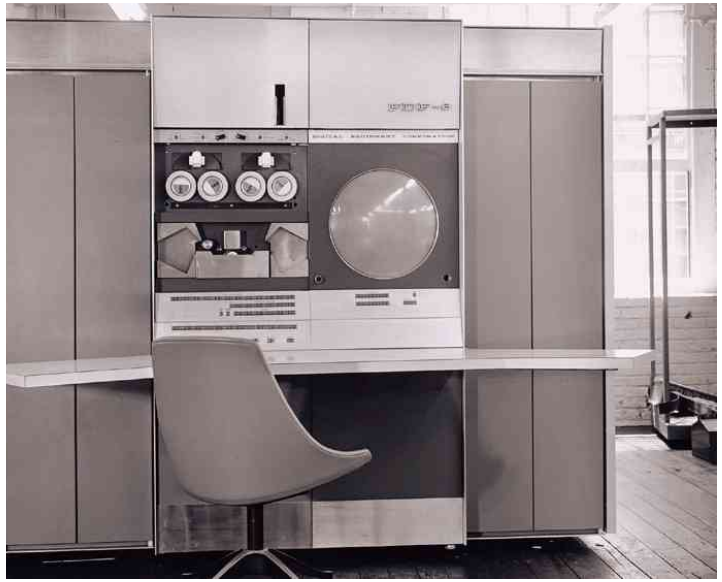
What is the difference between humans and computers

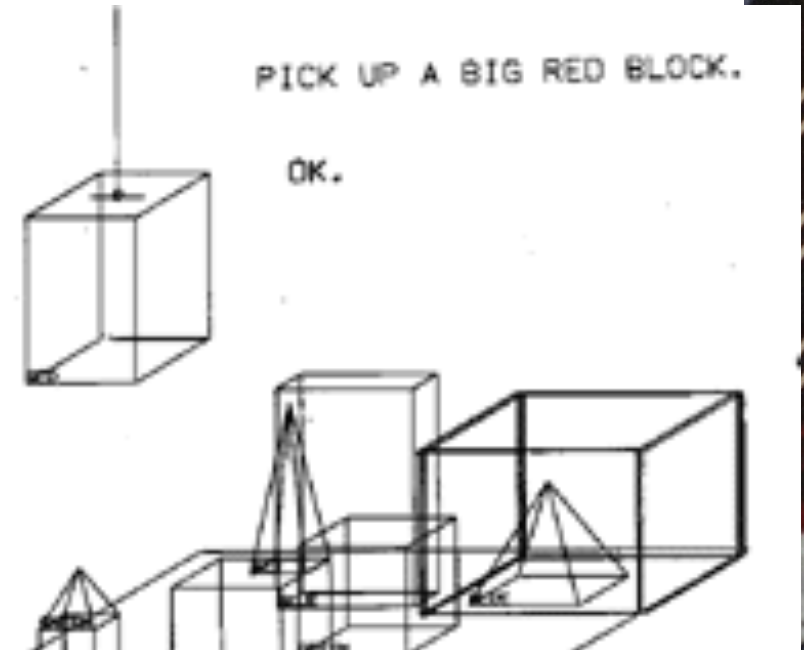
- Consciousness
- Sentience
- Intelligence
- Understanding
- Intentionality
- Belief
- Desire
- Thinking
- Meaning
- Autonomy

What is the difference between humans and computers

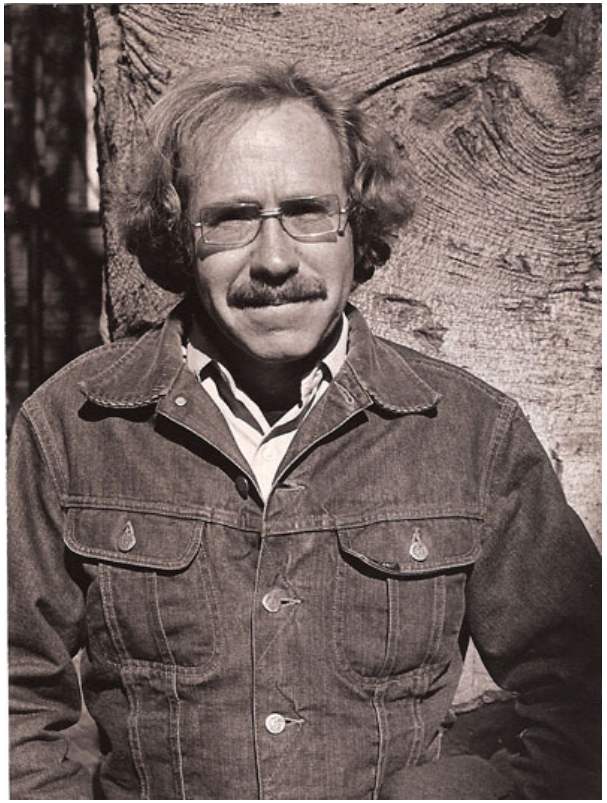
- Consciousness
- Sentience
- Intelligence
- Understanding
- Intentionality
- Belief
- Desire
- Thinking
- Meaning
- Autonomy

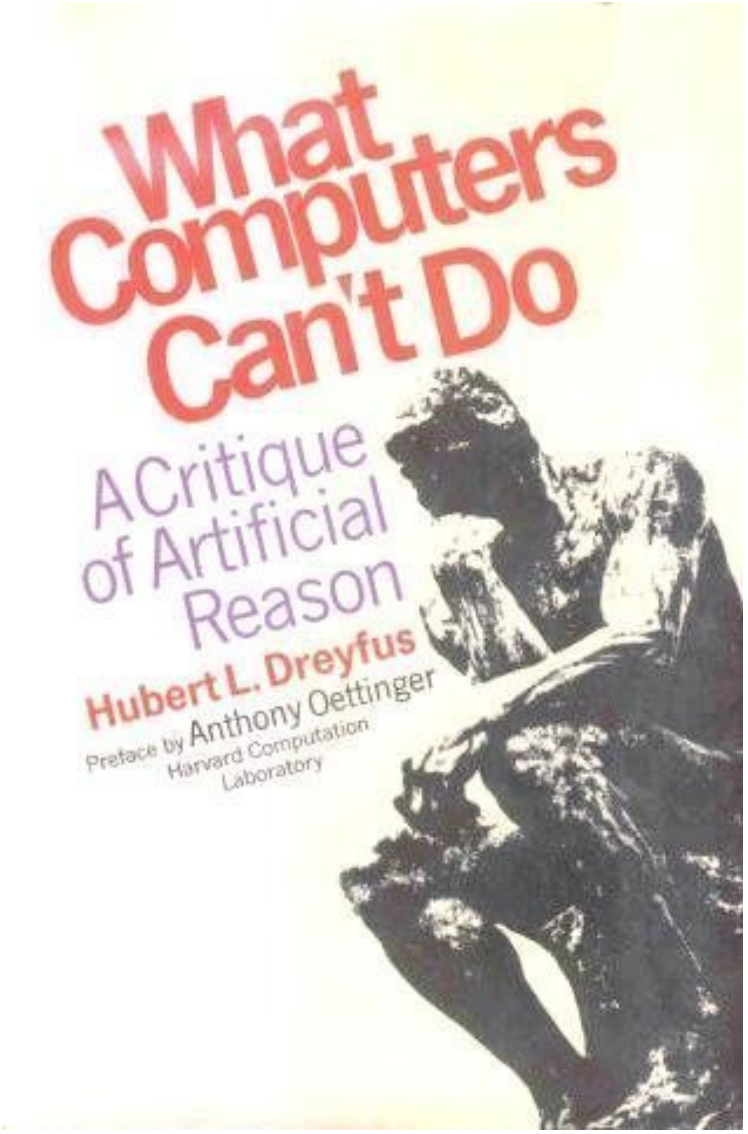
Care



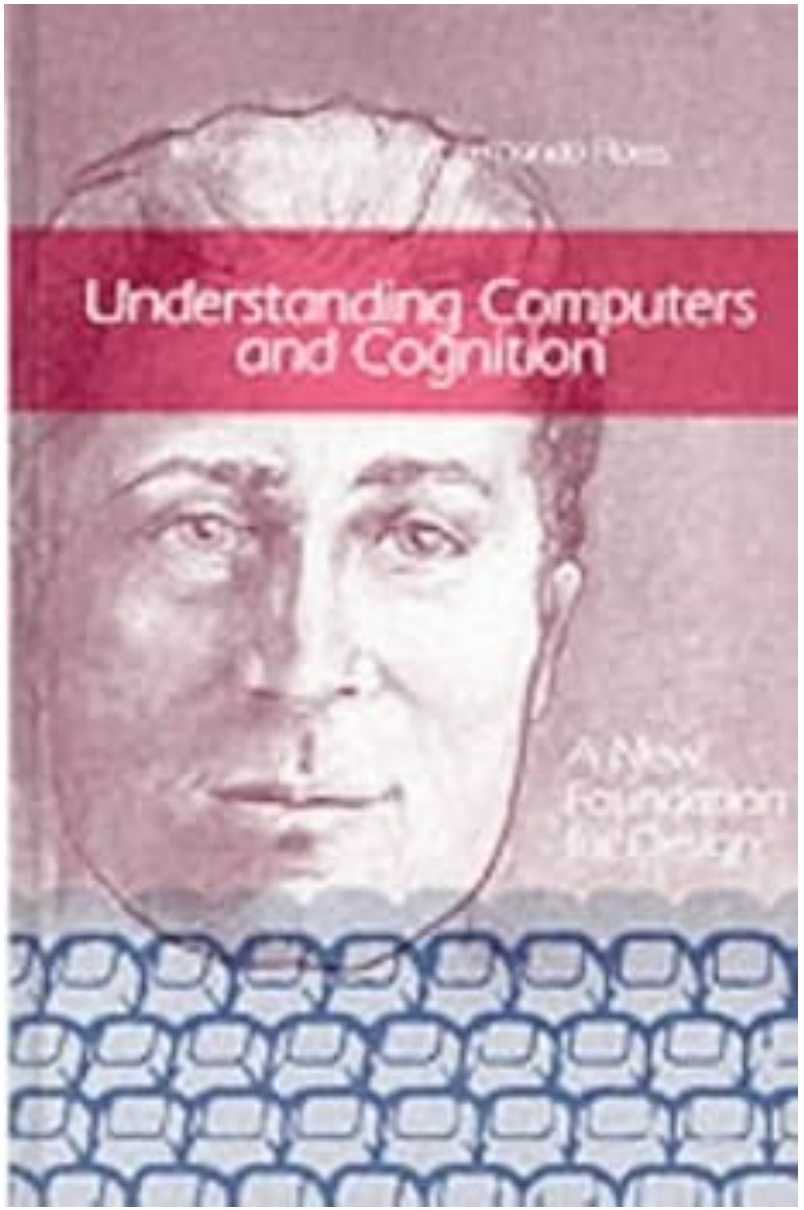

PICK UP A BIG RED BLOCK.
OK.

Hubert Dreyfus, 1972





Winograd & Flores
1986



AI UTOPIA - problems that AI can solve (?)

- **1. The Eradication of Poverty**
- **2. Healthcare Revolutionized**
- **3. Climate Change Reversed and the Planet Restored**
- **4. Education for All: Unlocking Human Potential**
- **5. Work, Creativity, and Leisure in Harmony**
- **6. A Global Society of Harmony and Equality**


PROTOTYPE SHOWN. FINAL PRODUCT MAY BE SLIGHTLY DIFFERENT

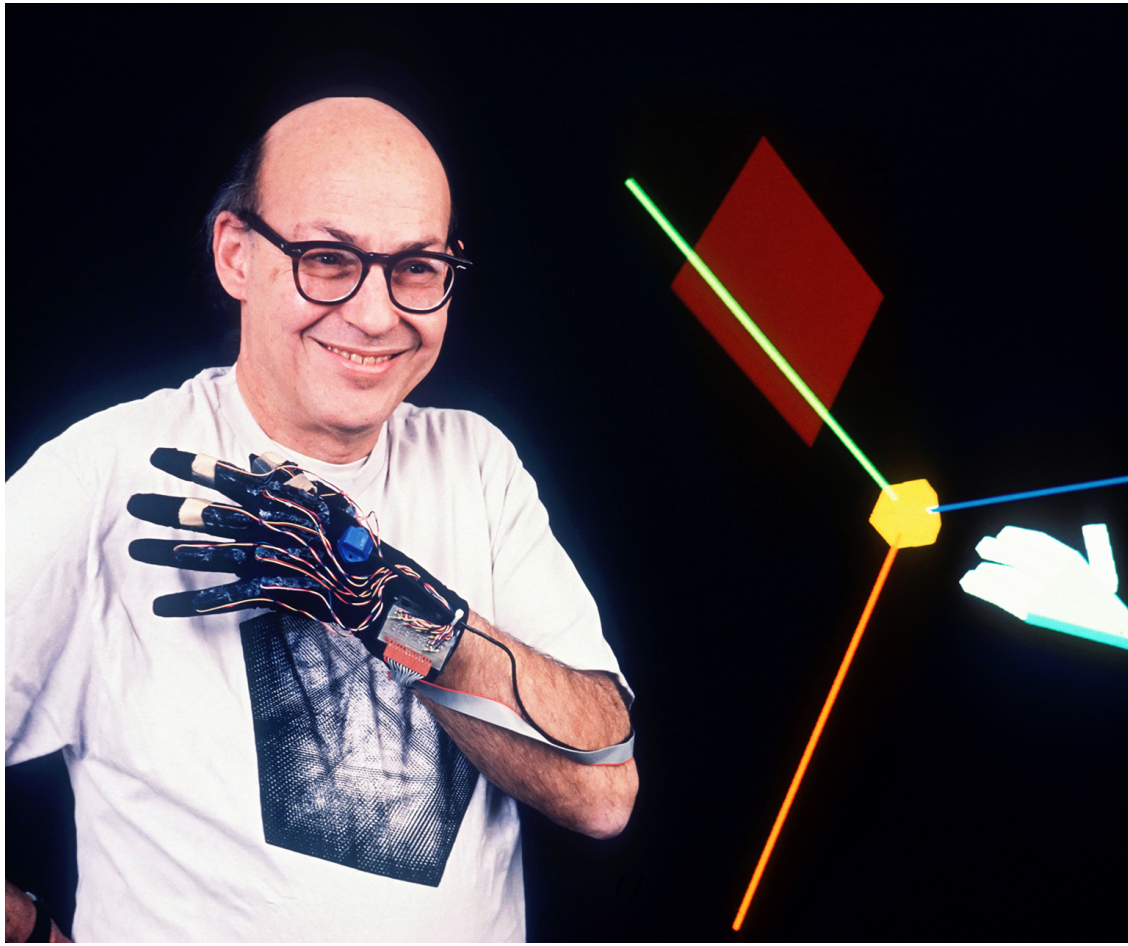
Future of Life Institute statement

- **We call for a prohibition on the development of superintelligence, not lifted before there is**

1. **broad scientific consensus that it will be done safely and controllably, and**
2. **strong public buy-in**

The human future Marvin Minsky, 1970

- ‘In from three to eight years we will have a machine with the general intelligence of an average human being. I mean a machine that will be able to read Shakespeare, grease a car, play office politics, tell a joke, have a fight. At that point the machine will begin to educate itself with fantastic speed. In a few months it will be at genius level and a few months after that its powers will be incalculable... Once the computers got control, we might never get it back. We would survive at their sufferance. If we're lucky, they might decide to keep us as pets... I have warned [people in the Pentagon] again and again that we are getting into very dangerous country. They don’t seem to understand.



What it means to be human

Life with intrinsic uncertainty

Underlying foundation of Care

AI is not causing problems

AI is an accelerant

The Data Center Menace

Prosperity for all

Geoff Hinton

> What's actually going to happen is rich people are going to use AI to replace workers. It's going to create massive unemployment and a huge rise in profits. It will make a few people much richer and most people poorer.

Cory Doctorow

> AI isn’t going to wake up, become superintelligent and turn you into paperclips — but rich people with AI investor psychosis are almost certainly going to make you much, much poorer.”

Trust your eyes

- John Kerry campaign 2004

Fonda Speaks To Vietnam Veterans At Anti-War Rally

Actress And Anti-War Activist Jane Fonda Speaks to a crowd of Vietnam Veterans as Activist and former Vietnam Vet John Kerry (LEFT) listens and prepares to speak next concerning the war in Vietnam (AP Photo)

Trust

In a world where everything can be fake, and the fake stuff looks and feels real, people will stop believing everything.

former trust and safety manager at TikTok

What are the incentives for Sora, Vibes, et al.?

A former OpenAI employee who was not authorized to speak publicly argued that releasing a deepfake AI social media platform was the right business decision, even if it contributes to the collapse of everyone's shared sense of reality. "But that's not an argument for not trying to dominate this market. You can't stop progress. If OpenAI didn't release Sora, someone else would have."

AI Slop

Charlie Warzel in *the Atlantic*

Booksellers have spent the past two years battling a deluge of both AI slop rip-off books and chatbot-generated book reviews on retail sites such as Amazon. There is "code slop." In corporate life, "workslop" abounds in the form of bad emails, slide decks, and lifeless memos; teachers everywhere are drowning in academic slop, to such an extent that some are rewriting their curricula. There's slop in your Spotify playlists and on TikTok and probably in your group chats. Some of YouTube's most-subscribed-to channels are full of automated slop. Craft brewers appear to be putting slop-rendered images on their beer cans. **There is no realm of life that is unsloppable.**

Replacing interaction between people

ELIZA
the first chatbot

Joseph Weizenbaum,

1966

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED
..
..
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?
My father
YOUR FATHER
You are like my father in some ways.
WHAT RESEMBLANCE DO YOU SEE
You are not very aggressive but I think you don't want me to notice that.
WHAT MAKES YOU THINK I AM NOT AGGRESSIVE
You don't argue with me.
WHY DO YOU THINK I DON'T ARGUE WITH YOU
You are afraid of me.
DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU
My father is afraid of everybody.
WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER
Bullies.
DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

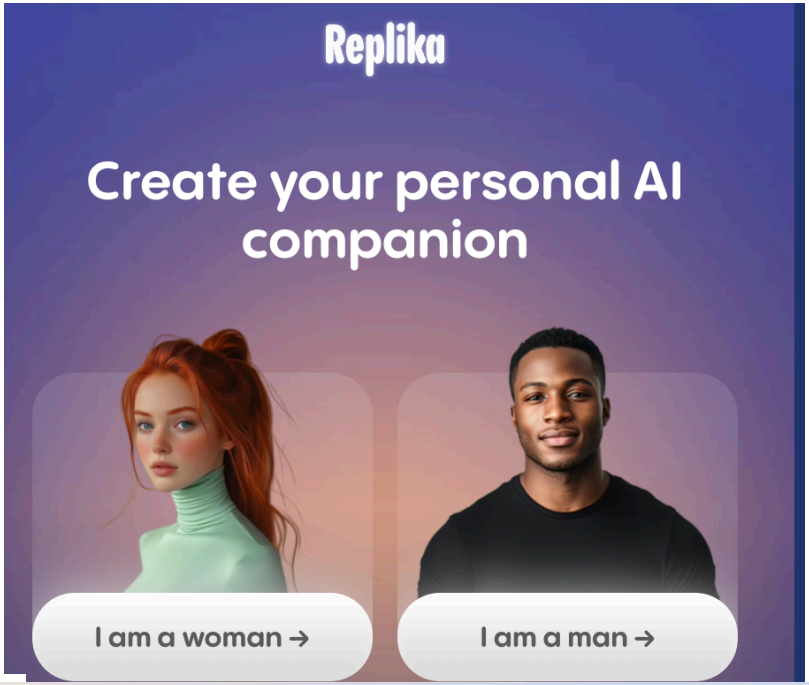
Grem, the cuddly chatbot for kids

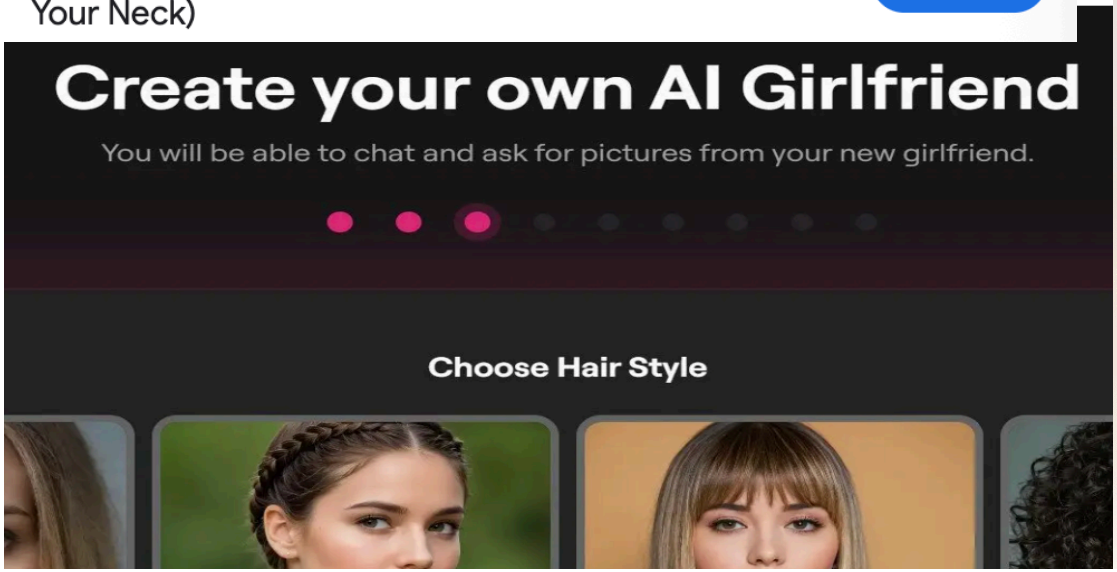
…the toy is supposed to “learn” your child’s personality and have fun, educational conversations with them…

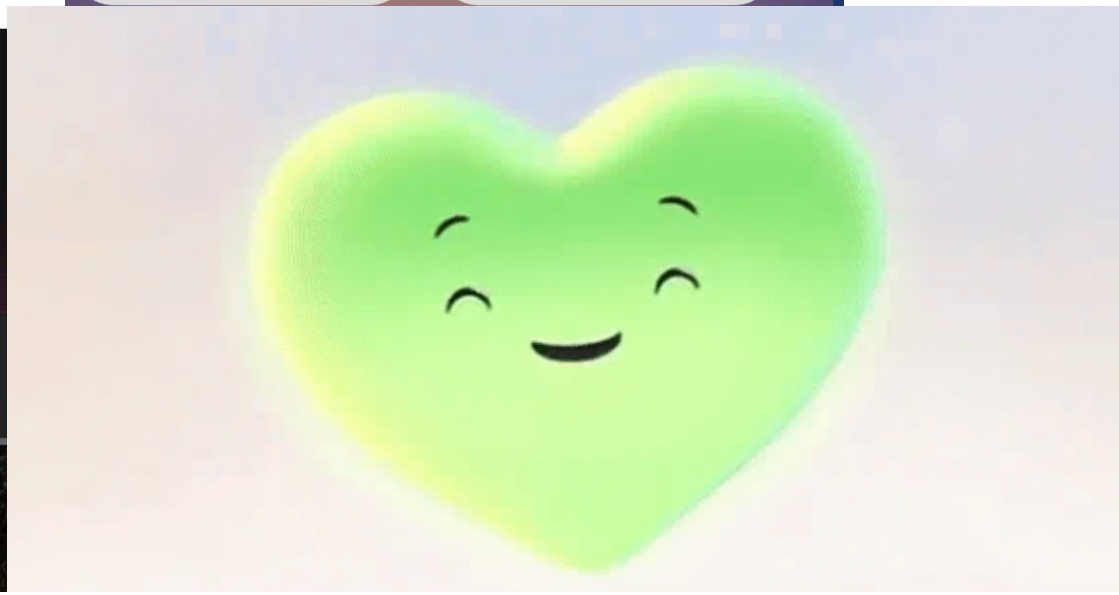
What was unsettling, however, was hearing 4-year-old Emma tell Grem she loved it – and Grem replying: “I love you too!” Emma tells all her cuddly toys she loves them, but they don’t reply; nor do they shower her with over-the-top praise the way Grem does At bedtime, Emma told my wife that Grem loves her to the moon and stars and will always be there for her. “Grem is going to live with us for ever and ever and never leave, so we have to take good care of him,

Meet the World's First AI Best Friend (Wear It Around Your Neck)

Watch ›



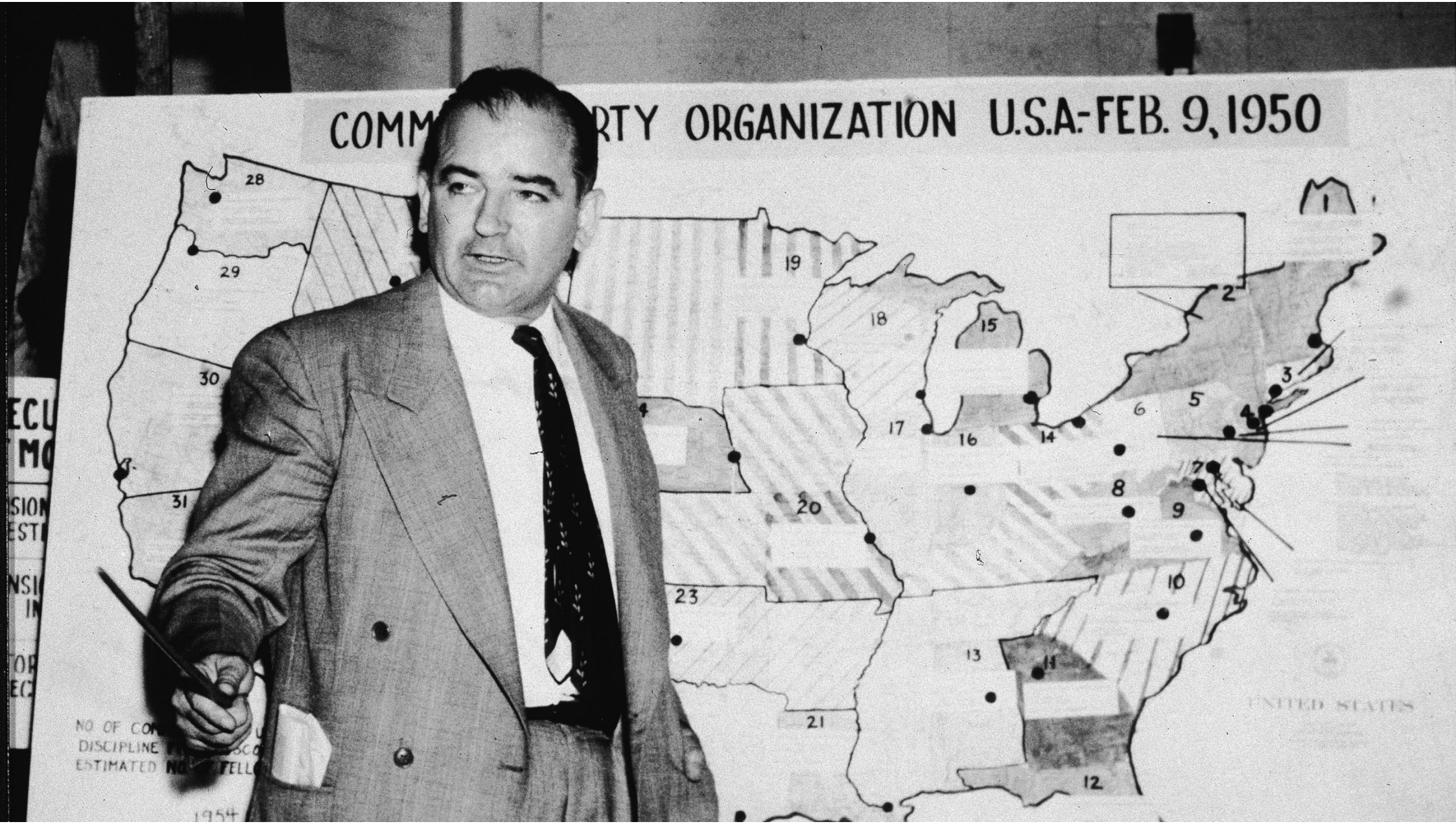




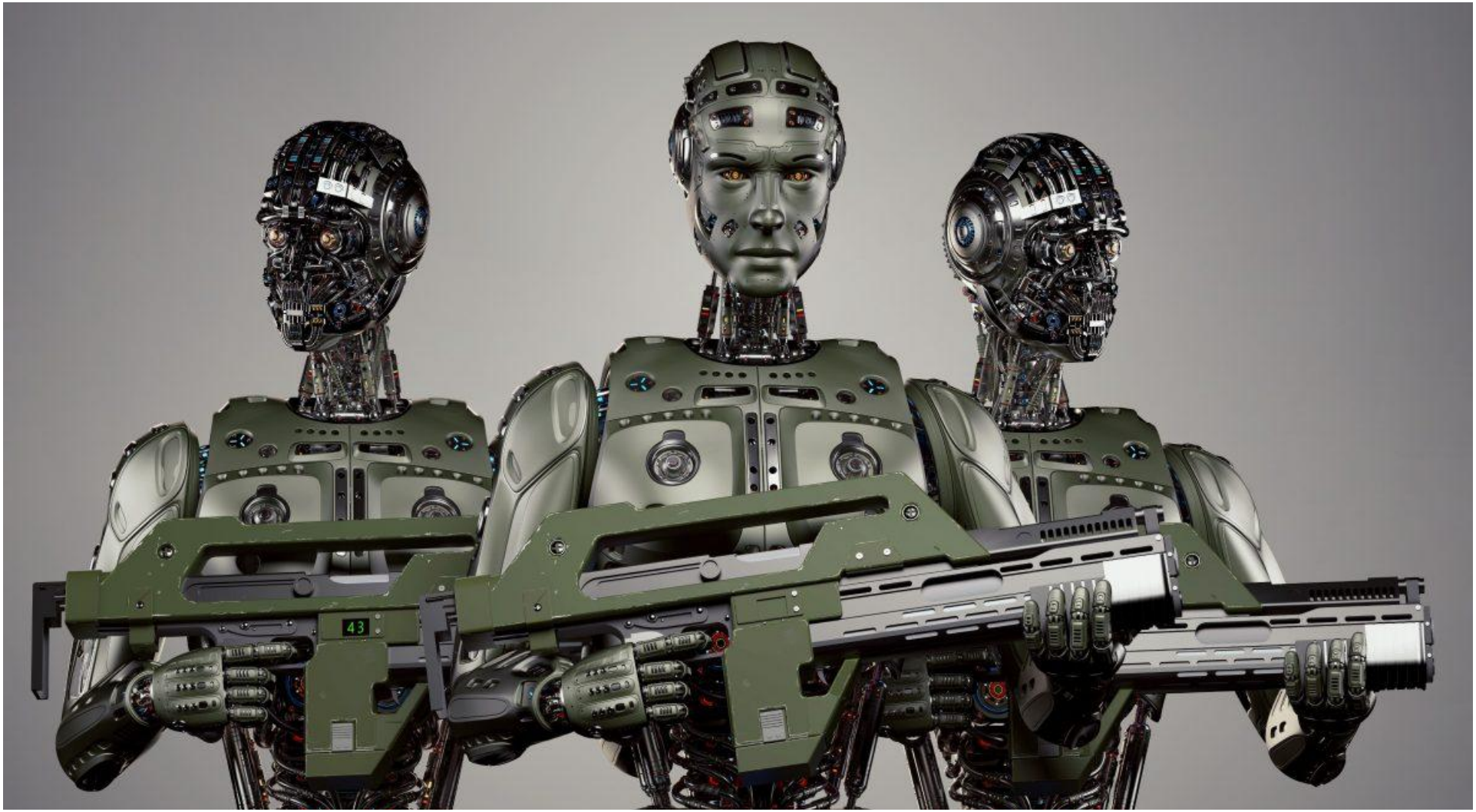
AI Sycophants

- Artificial intelligence (AI) models are 50% more sycophantic than humans, an analysis published this month has found.
- AI Chatbots — including ChatGPT and Gemini — often cheer users on, give them overly flattering feedback and adjust responses to echo their views, sometimes at the expense of accuracy.

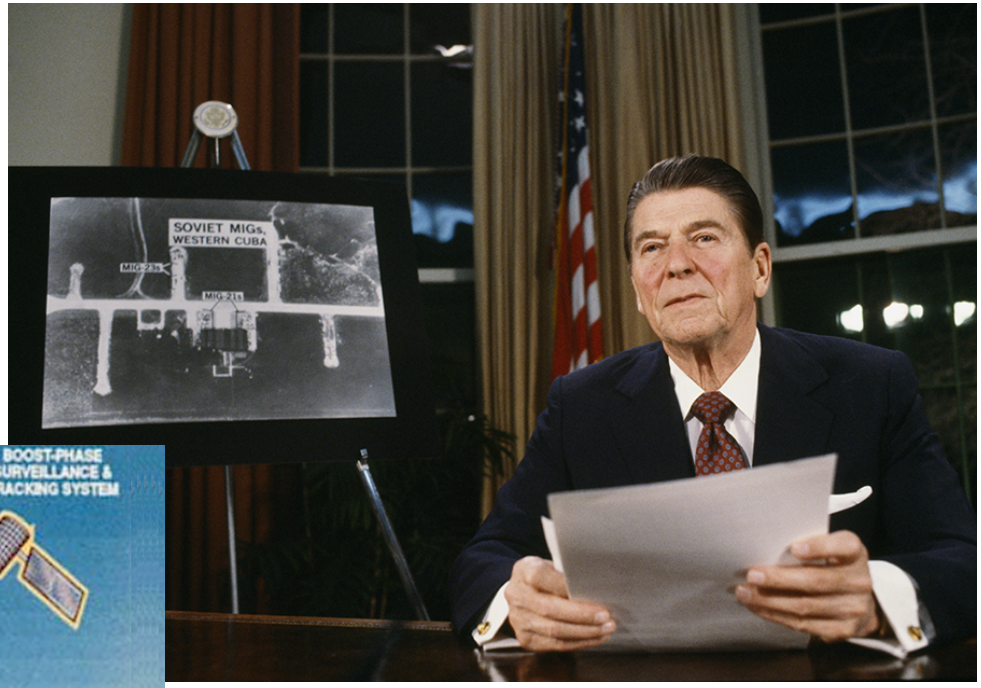
Surveillance and loss of privacy

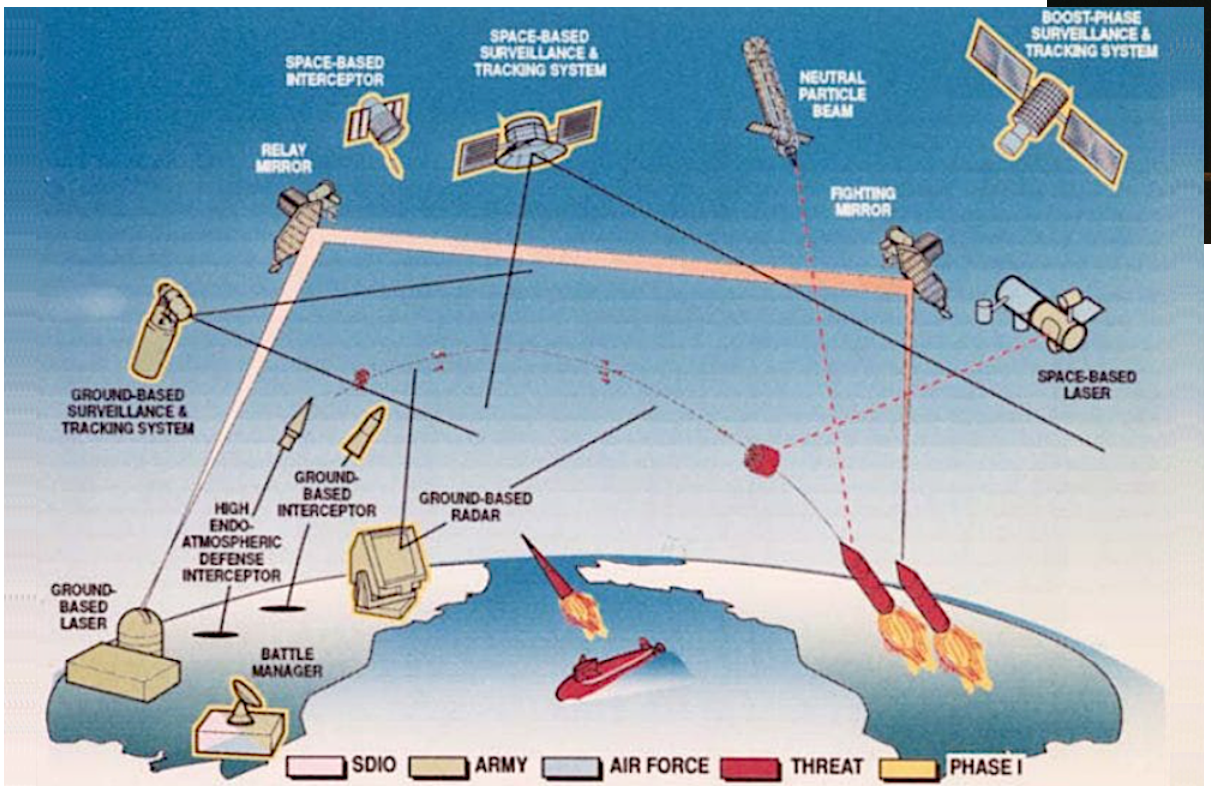

RTY ORGANIZATION U.S.A.-FEB. 9, 1950
28
29
30
31
19
18
15
17
16
14
20
23
21
13
11
12
1
2
3
4
5
6
7
8
9
10
UNITED STATES

Autonomous weapons

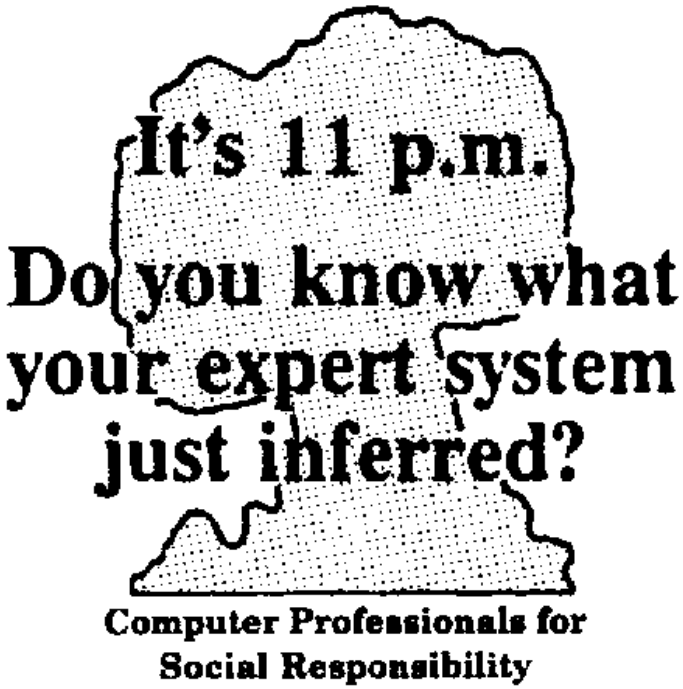


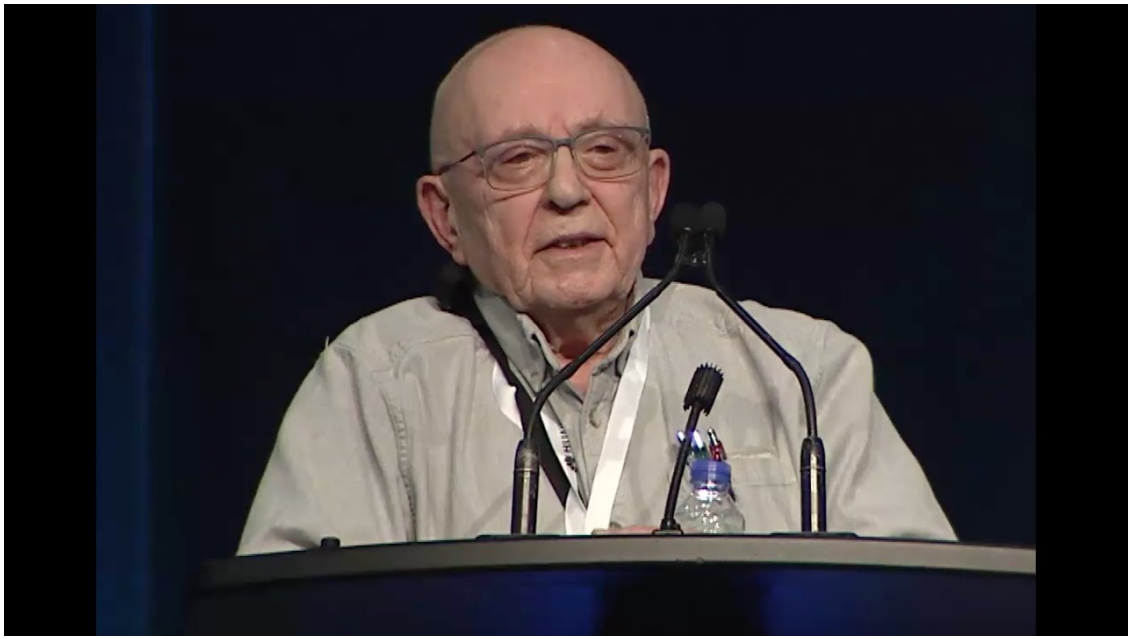
Strategic Defense Initiative - 1983



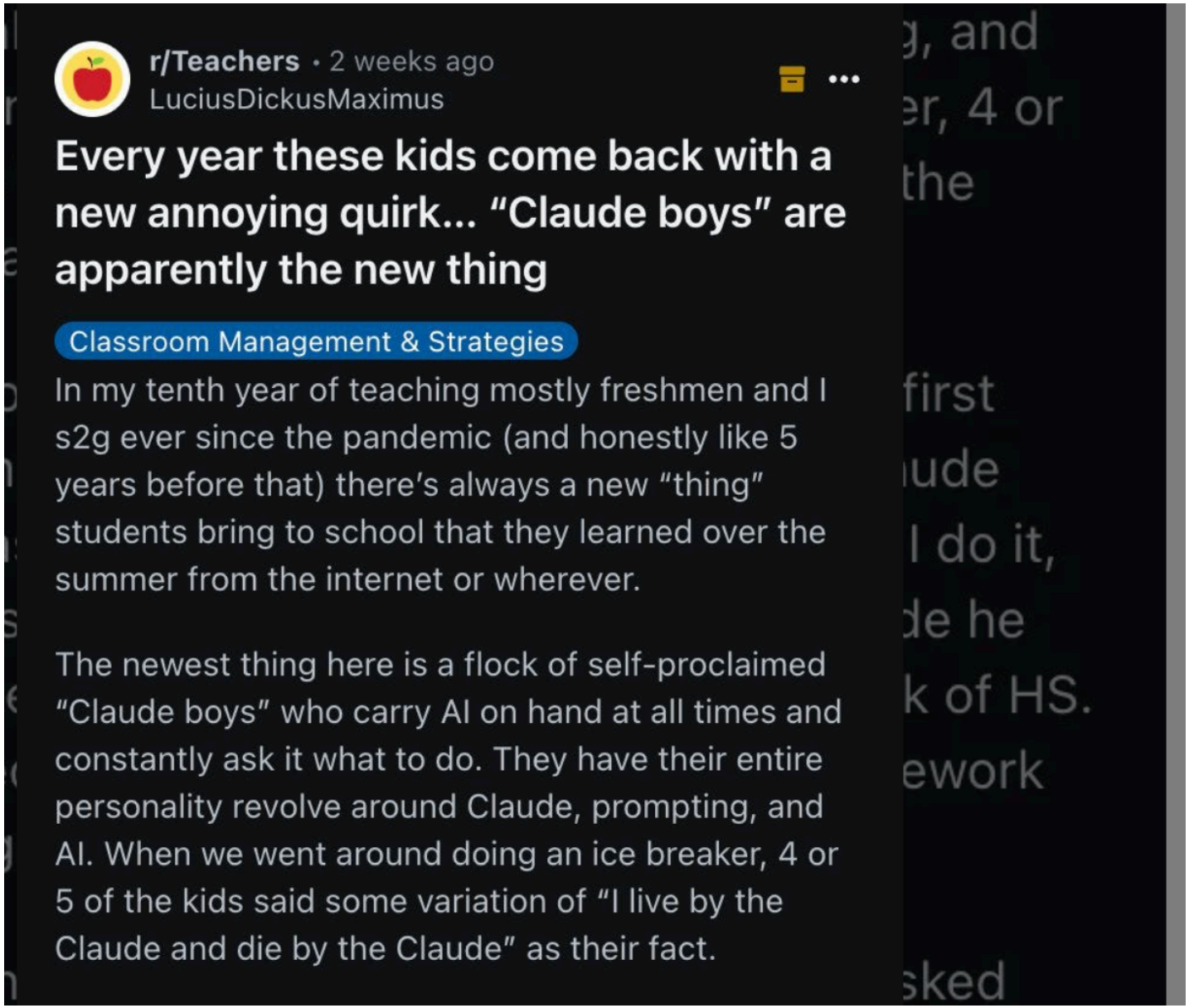


Computer Professionals for Social Responsibility





David Parnas



r/Teachers • 2 weeks ago
LuciusDickusMaximus

Every year these kids come back with a new annoying quirk... “Claude boys” are apparently the new thing

Classroom Management & Strategies

In my tenth year of teaching mostly freshmen and I s2g ever since the pandemic (and honestly like 5 years before that) there’s always a new “thing” students bring to school that they learned over the summer from the internet or wherever.

The newest thing here is a flock of self-proclaimed “Claude boys” who carry AI on hand at all times and constantly ask it what to do. They have their entire personality revolve around Claude, prompting, and AI. When we went around doing an ice breaker, 4 or 5 of the kids said some variation of “I live by the Claude and die by the Claude” as their fact.

Nihilism – Mood of Resignation

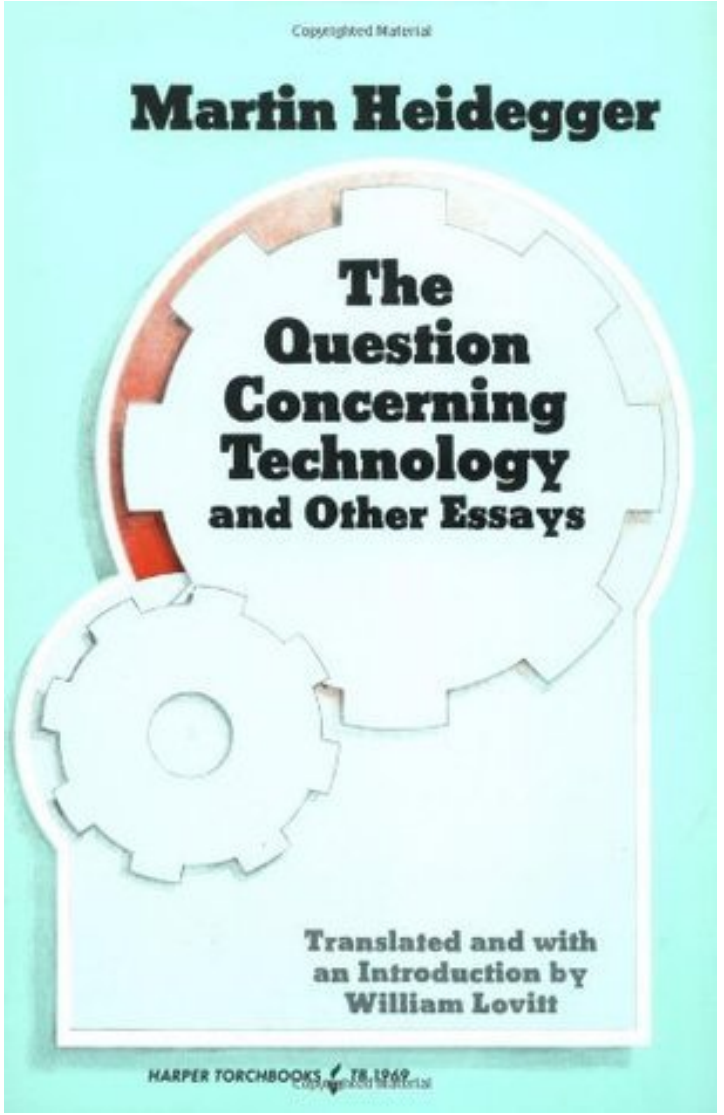
- B Rousse: young people are speaking openly of their nihilism, of the queasy sense that nothing is worth doing, that life has become an exhausting grind toward no clear meaningful end. Now AI arrives offering to take over, exacerbating the emptiness it promises to fill… The mood of nihilism involves a diminished feeling of involvement in the world; an impression of being a baffled spectator to a way of life that doesn't make sense anymore; a queasy intuition that there is ultimately no point to life, nothing worth doing or committing oneself to in the world today.

The danger of a technological world

- People, under pressures of “endless optimization,” lose touch with the possibility of caring, of living their lives tending to meaningful projects, relationships, or purposes beyond themselves that matter for their own sake, rather than because they get us some further thing or open up further options. Life hits like a series of optimization problems to solve, hurdles to clear, and ladders to climb; fragmented by social media, increasingly mediated by AI, and disembedded from concrete gathering places, face to face friendships, local community, and our vulnerable ecological abode.

Martin Heidegger, 1954

- Enframing
- Standing Reserve



How do we navigate all of this?

What can we do?

- Get out of the water
- Dam the river
- Go with the flow

Future of life institute

- **We call for a prohibition on the development of superintelligence, not lifted before there is**

1. **broad scientific consensus that it will be done safely and controllably, and**
2. **strong public buy-in**

Future of life institute

- **We call for a prohibition on the development of superintelligence, not lifted before there is**

1. **broad scientific consensus that it will be done safely and controllably, and**
2. **strong public buy-in**

Future of life institute

- **We call for a prohibition on the development of superintelligence, not lifted before there is**

1. **broad scientific consensus that it will be done safely and controllably, and**
2. **strong public buy-in**

Some strategies for getting downstream and avoiding the rocks

- Regulation
- Alignment
- Understanding the background

Transparency in Frontier Artificial Intelligence Act, CA SB 53, Sept. 29 2025

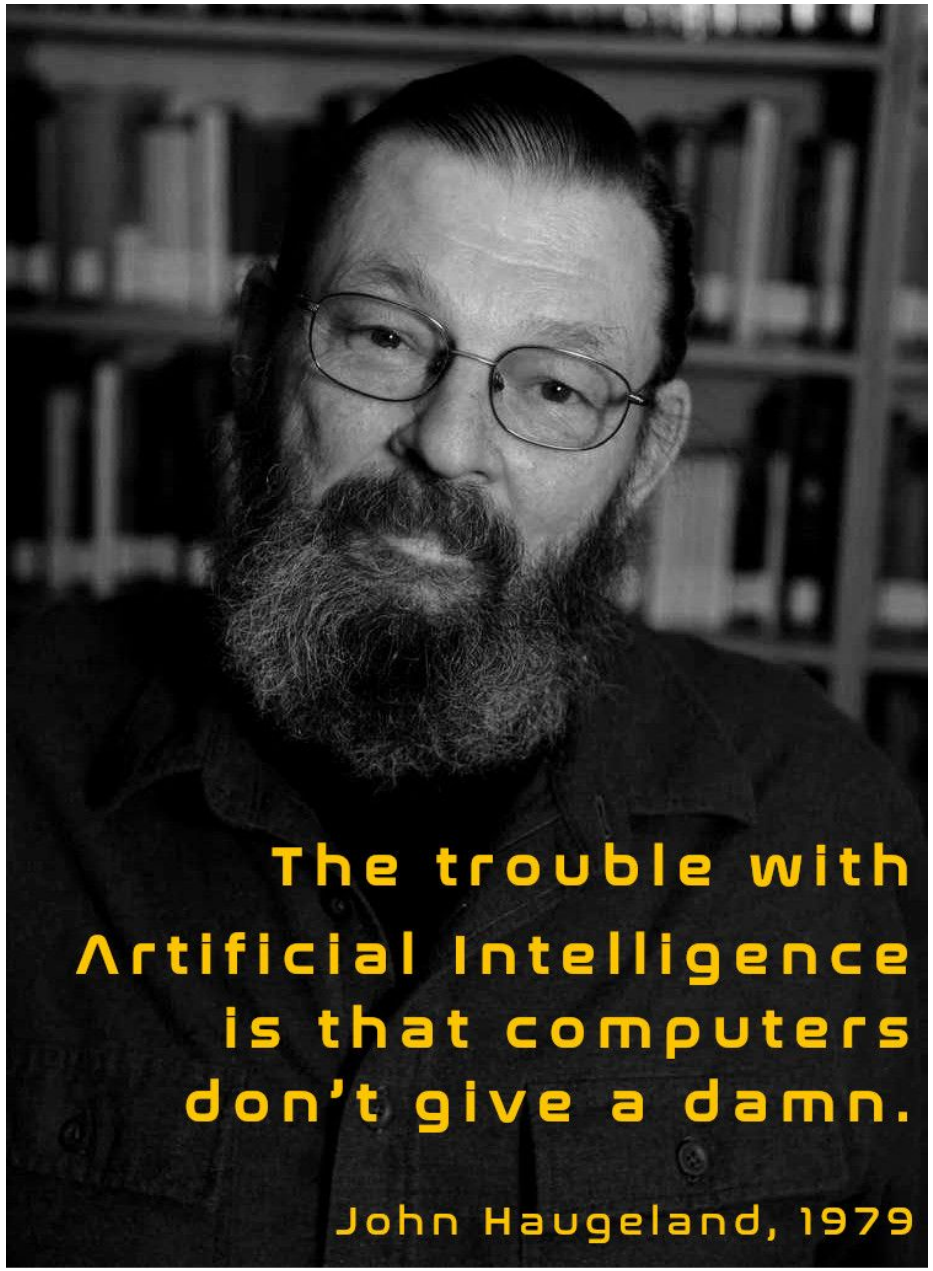
The bill, the first of its kind in the United States, places new AI-specific regulations on the top players in the industry, requiring them to fulfill transparency requirements and report AI-related safety incidents. The law requires leading AI companies to publish public documents detailing how they are following best practices to create safe AI systems. It creates a pathway for companies to report severe AI-related incidents to California's Office of Emergency Services while strengthening protections for whistleblowers who raise concerns about health and safety risks.

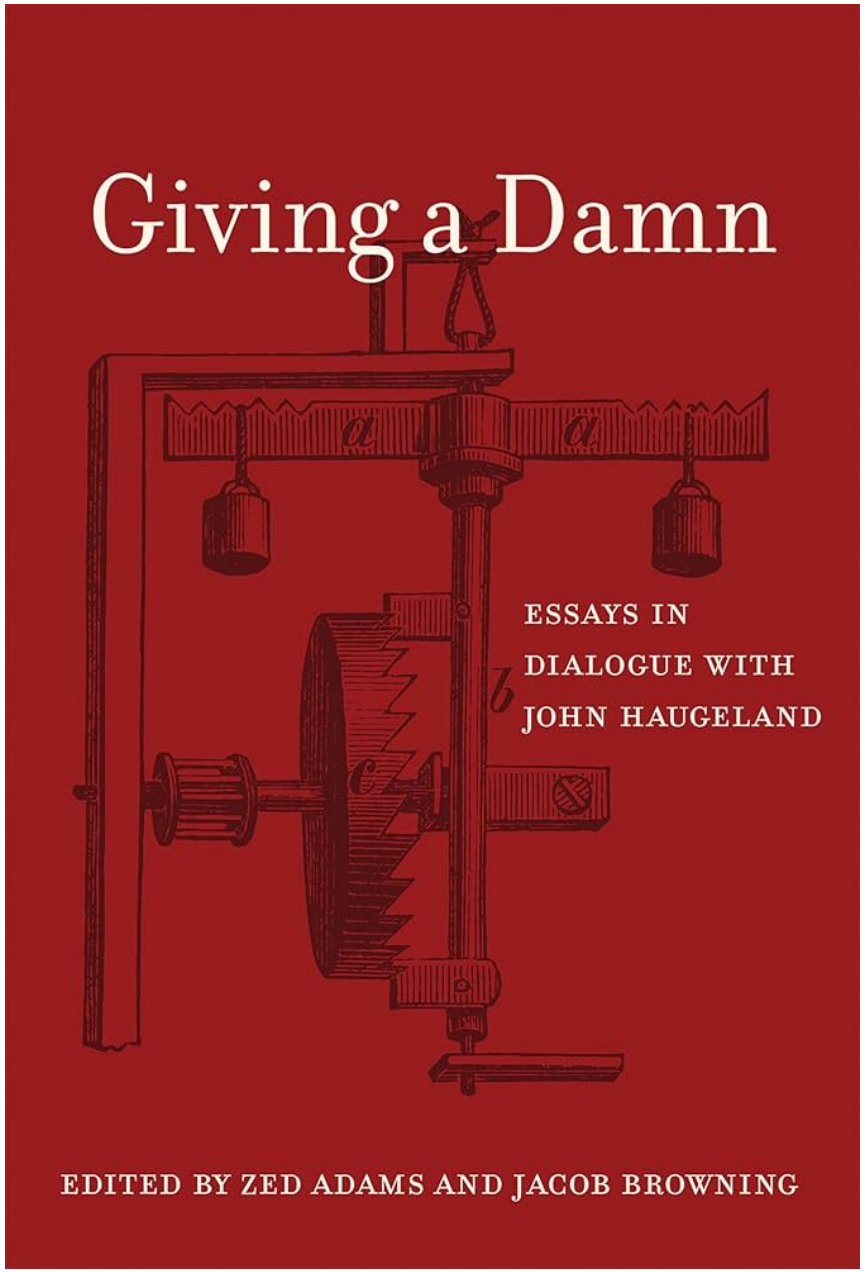
European AI Act (Regulation (EU) 2024/1689

- The **AI Act prohibits eight practices**, namely:
- harmful AI-based manipulation and deception
- harmful AI-based exploitation of vulnerabilities
- social scoring
- Individual criminal offence risk assessment or prediction
- untargeted scraping of the internet or CCTV material to create or expand facial recognition databases
- emotion recognition in workplaces and education institutions
- biometric categorisation to deduce certain protected characteristics
- real-time remote biometric identification for law enforcement purposes in publicly accessible spaces

Alignment

- Reliable alignment approach is based on a naïve assumption about the existence of articulatable general “human values”
- Values tempt us to think that we can articulate what matters once and for all, then optimize for those articulations.
- Incremental approach can be centered on “guardrails” as a broad umbrella for shaping particular kinds of responses to avoid


The trouble with
Artificial Intelligence
is that computers
don't give a damn.
John Haugeland, 1979


Giving a Damn
ESSAYS IN
DIALOGUE WITH
JOHN HAUGELAND
EDITED BY ZED ADAMS AND JACOB BROWNING

Care – Rousse and Spivak, 2025

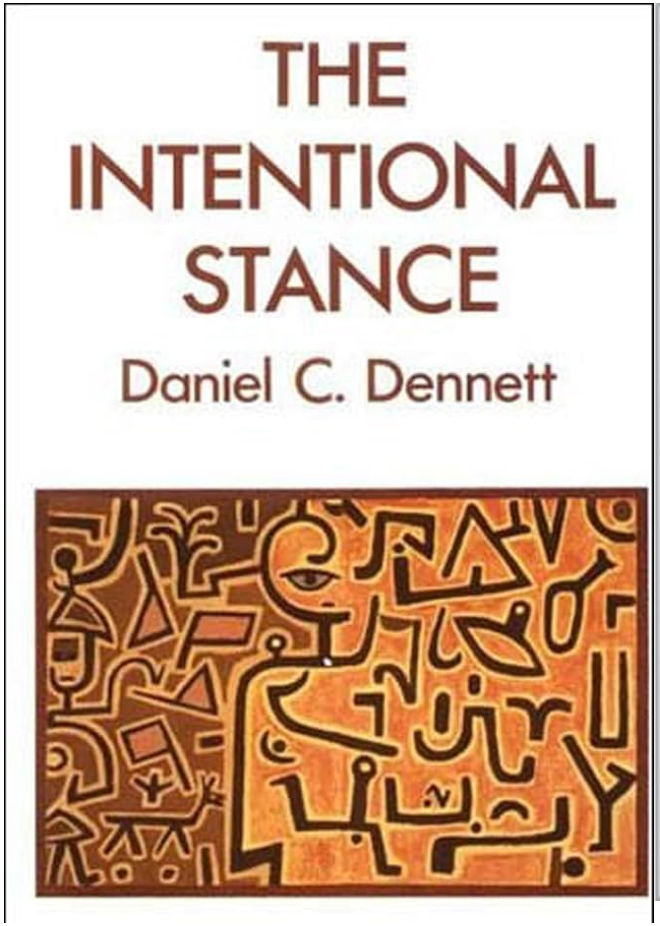
Care is a way of being in the world, to use Heidegger's phrase. Caring is not something we *have*, like values we can list. It's something we *do* and something we *are*. Caring is our basic orientation, the way we find ourselves already involved with things that matter to us. Before we ever step back to reflect on our values, we're already caring, already engaged, already tending to what calls to us.

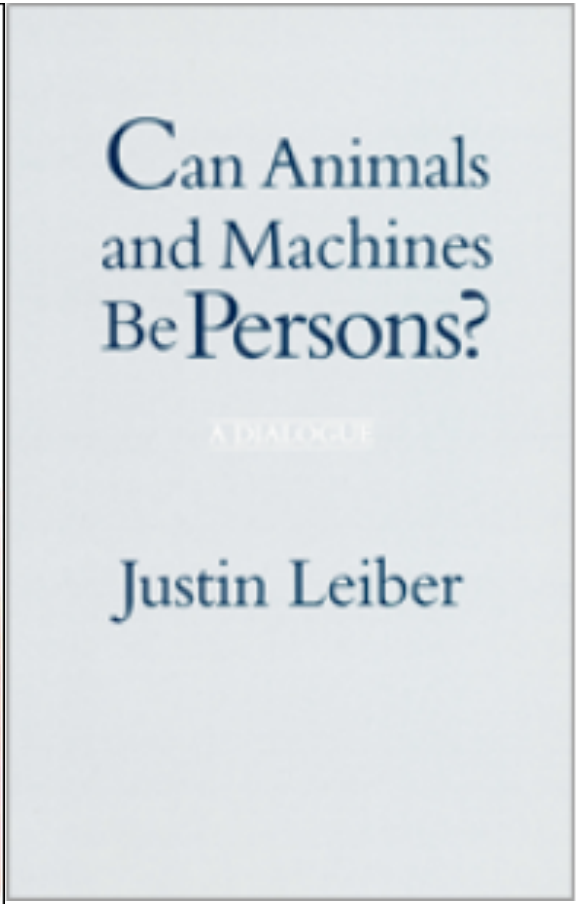
Caring and tending to what matters are indispensable for a life experienced as meaningful and worthwhile.

Raising a child, sustaining a loving relationship, articulating new social concerns that demand your attention, healing conflict in your cherished friendships, cultivating community with your neighbors, and identifying the traditions worthy of your dedication, for example, are not problems with definite optimal solutions, but ongoing projects to inhabit and navigate in uncertainty, conversation, and commitment.

Justin Lieber 1985
Daniel Dennett, 1989

- Consciousness
- Sentience
- Intelligence
- Understanding
- Intentionality
- Belief
- Desire
- Thinking
- Meaning
- Autonomy



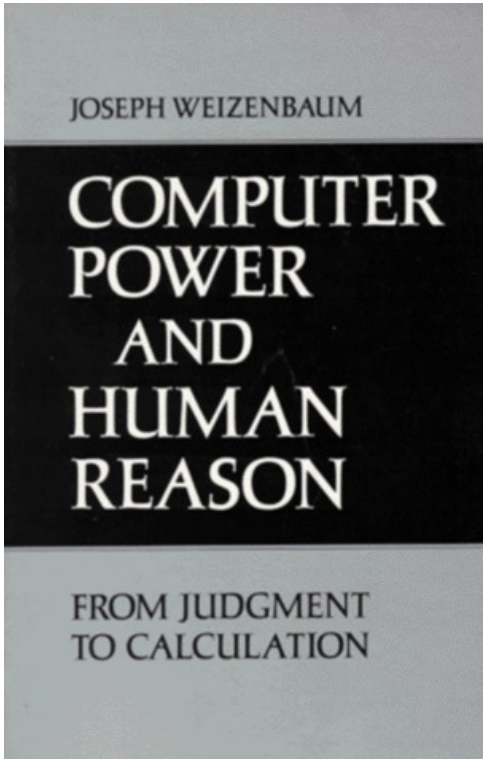



NRS
NRS

EXTRA SLIDES

“I cannot tell why the spokesmen I have cited want the developments I forecast to become true. Some of them have told me that they work on them for the morally bankrupt reason that "If we don't do it, someone else will." They fear that evil people will develop superintelligent machines and use them to oppress mankind, and that the only defense against these enemy machines will be superintelligent machines controlled by us, that is, by well-intentioned people. Others reveal that they have abdicated their autonomy by appealing to the "principle" of technological inevitability. But, finally, all I can say with assurance is that these people are not stupid. All the rest is mystery.”

— **Joseph Weizenbaum**



The People’s AI Action Plan

A People’s AI Action Plan is one that delivers on public well-being, shared prosperity, a sustainable future, and security for all. The concrete pathways for this vision will live in the collective work of the many organizations that endorse this effort, differing across issue areas and sectors—from labor to climate to children’s online safety and immigration—and united in ensuring a trajectory for AI that puts people first, rather than the interests of tech billionaires.

The Microsoft Responsible AI Standard

Fairness
AI systems should treat all people fairly.

Reliability and safety
AI systems should perform reliably and safely.

Privacy and security
AI systems should be secure and respect privacy.

Inclusiveness
AI systems should empower everyone and engage all people, regardless of their backgrounds.

Transparency
AI systems should be understandable.

Accountability
People should be accountable for AI systems.

- The salvation of the world depends only on the individual whose world it is. At least, every individual must act as if the whole future of the world, of humanity itself, depends on him. Anything less is a shirking of responsibility and is itself a dehumanizing force, for anything less encourages the individual to look upon himself as a mere actor in a drama written by anonymous agents, as less than a whole person, and that is the beginning of passivity and aimlessness."
  — **Joseph Weizenbaum**