biocomm AI
SAFE AI BLOG

Blog
Resources
Categories

Homo sapiens vs. Machine intelligence
SV vs. CA
The gloves are off! The fight for the future of humanity begins.
[Hint: WE are in control, for the moment.]

SV vs. CA. Homo sapiens vs. Machine intelligence. The gloves are off! The fight for the future of humanity begins.

June 10, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

SCOTT WIENER
REPRESENTING SENATE DISTRICT 11

PRESS RELEASE

In A Bipartisan Vote, Senate Passes Senator Wiener's Landmark AI Safety And Innovation Bill

[SV asked lawmakers for regulation. Now they get it from State of CA]

PRESS RELEASE. In A Bipartisan Vote, Senate Passes Senator Wiener's Landmark AI Safety And Innovation Bill . MAY 21, 2024.

June 10, 2024

Silicon Valley (SV) literally begged for regulation from lawmakers. [...]

Financial Times
Silicon Valley in uproar over Californian AI safety bill.
[but they did ask lawmakers for regulation]

FT. Silicon Valley in uproar over Californian AI safety bill. Tech companies launch fightback against proposed law to introduce 'kill switch' on powerful artificial intelligence models.

June 10, 2024

A very important read from an extremely respected source [...]

SITUATIONAL AWARENESS: The Decade Ahead

Leopold Aschenbrenner, June 2024

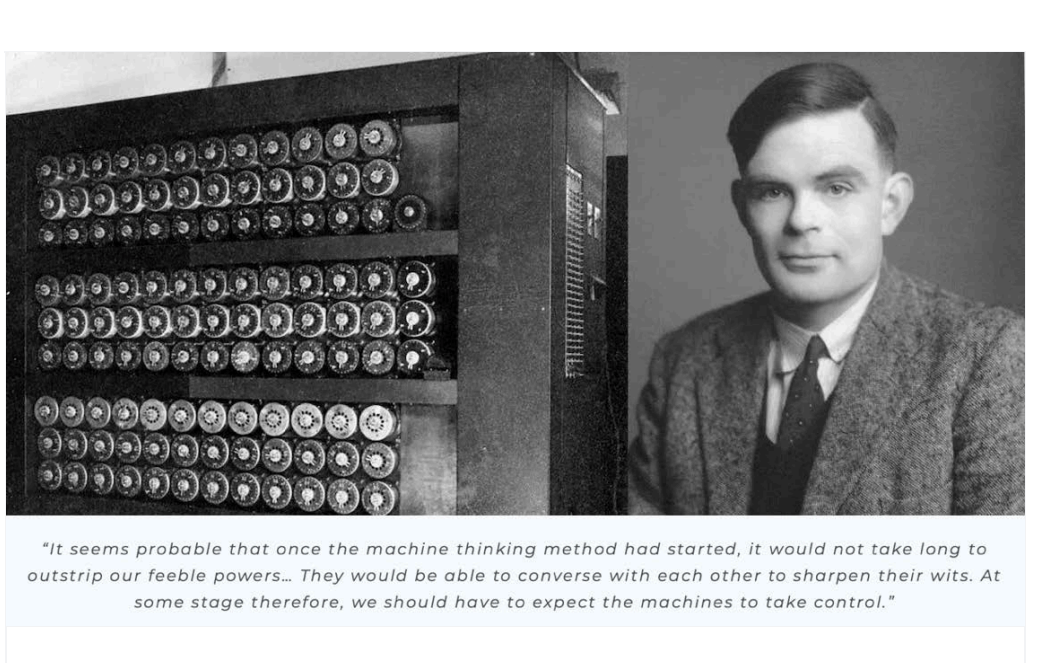
You can see the future first in San Francisco.

Over the past year, the talk of the town has shifted from $10 billion compute clusters to $100 billion clusters to trillion-dollar clusters. Every six months another zero is added to the boardroom plans.

SITUATIONAL AWARENESS: The Decade Ahead. Leopold Aschenbrenner, June 2024.

June 9, 2024

A very good read from a respected source! Be [...]


"It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers... They would be able to converse with each other to sharpen their wits. At some stage therefore, we should have to expect the machines to take control."

Alan Turing (1951). "...we should have to expect the machines to take control."

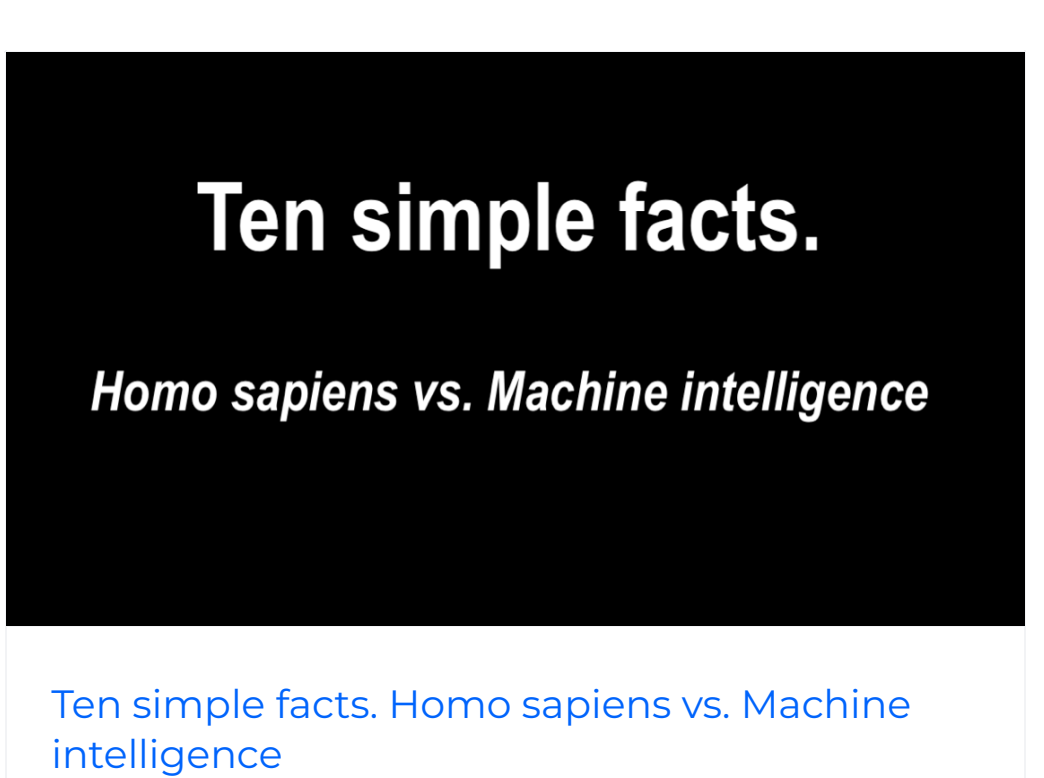
June 8, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

OLMo. The Allen Institute for Artificial Intelligence. Open Language Model: OLMo A State-of-the-Art, Truly Open LLM and Framework.

June 4, 2024

WOW. Be VERY careful what you wish for! [...]


Ten simple facts.
Homo sapiens vs. Machine intelligence

Ten simple facts. Homo sapiens vs. Machine intelligence

June 1, 2024

Be VERY careful what you wish for! [...]

Science

Managing extreme AI risks amid rapid progress

Preparation requires technical research and development, as well as adaptive, proactive governance

SCIENCE. Managing extreme AI risks amid rapid progress Preparation requires technical research and development, as well as adaptive, proactive governance. 20 MAY.

May 31, 2024

A very important read from extremely respected and knowledgable [...]

They Make Me Sick - SafeAI Forever
Safe FOREVER AI
THEY MAKE ME SICK!
Watch on YouTube

"They Make Me Sick". So SIMPLE- even A CHILD can Understand the Existential Danger of AI (in 75 seconds)

May 28, 2024

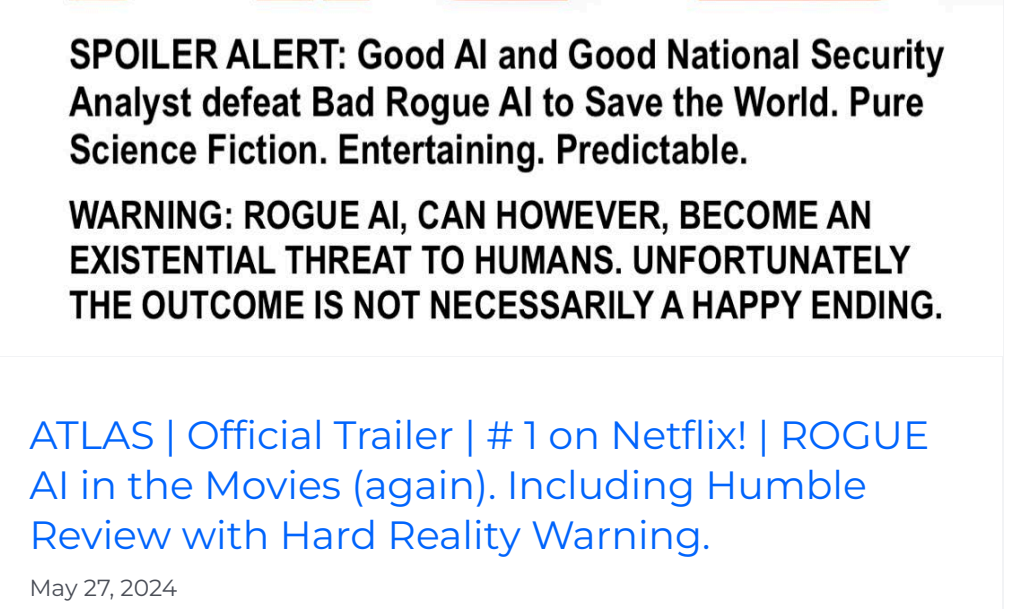
SO SIMPLE! Even a CHILD can Understand the Existential [...]

China's military shows off rifle-toting robot dogs
CNN

CNN. China's military shows off rifle-toting robot dogs. 28 MAY.

May 28, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


N TRAILER STARRING JENNIFER LOPEZ
ATLAS
SPOILER ALERT: Good AI and Good National Security Analyst defeat Bad Rogue AI to Save the World. Pure Science Fiction. Entertaining. Predictable.
WARNING: ROGUE AI, CAN HOWEVER, BECOME AN EXISTENTIAL THREAT TO HUMANS. UNFORTUNATELY THE OUTCOME IS NOT NECESSARILY A HAPPY ENDING.

ATLAS | Official Trailer | # 1 on Netflix! | ROGUE AI in the Movies (again). Including Humble Review with Hard Reality Warning.

May 27, 2024

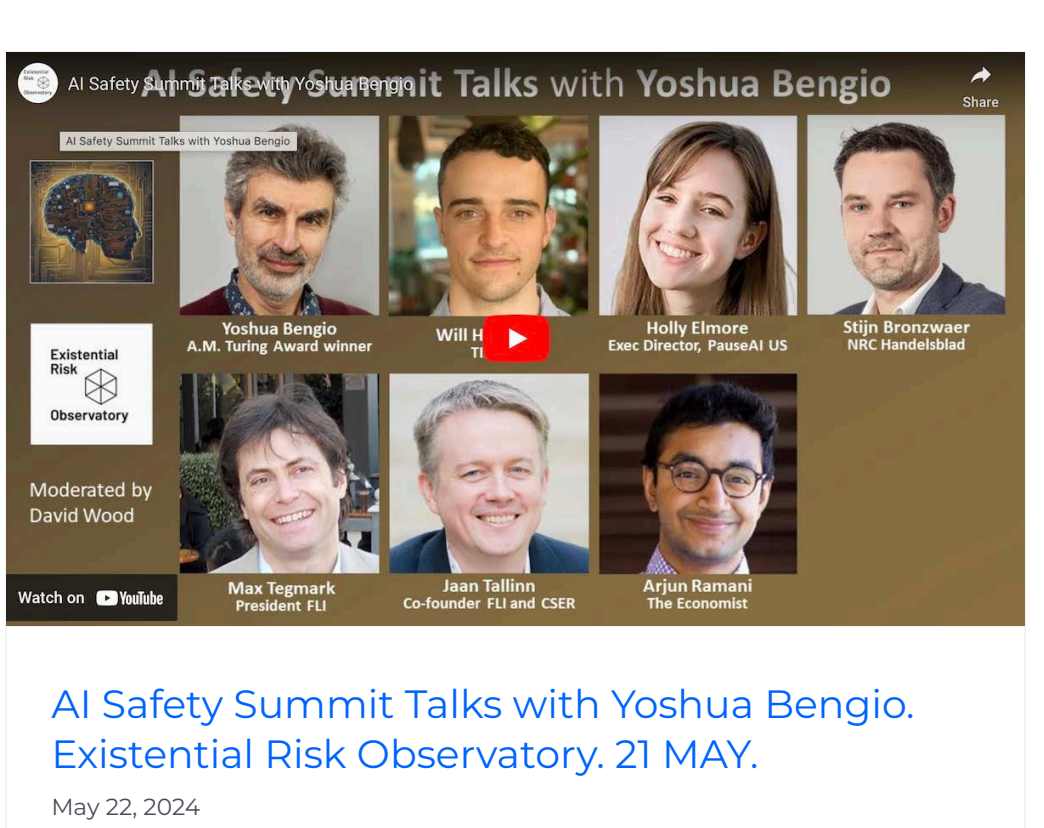
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Max Tegmark:
"Big tech has distracted world from existential risk of AI"
The Guardian

THE GUARDIAN. Big tech has distracted world from existential risk of AI, says top scientist. Max Tegmark argues that the downplaying is not accidental and threatens to delay, until it's too late, the strict regulations needed.

May 25, 2024

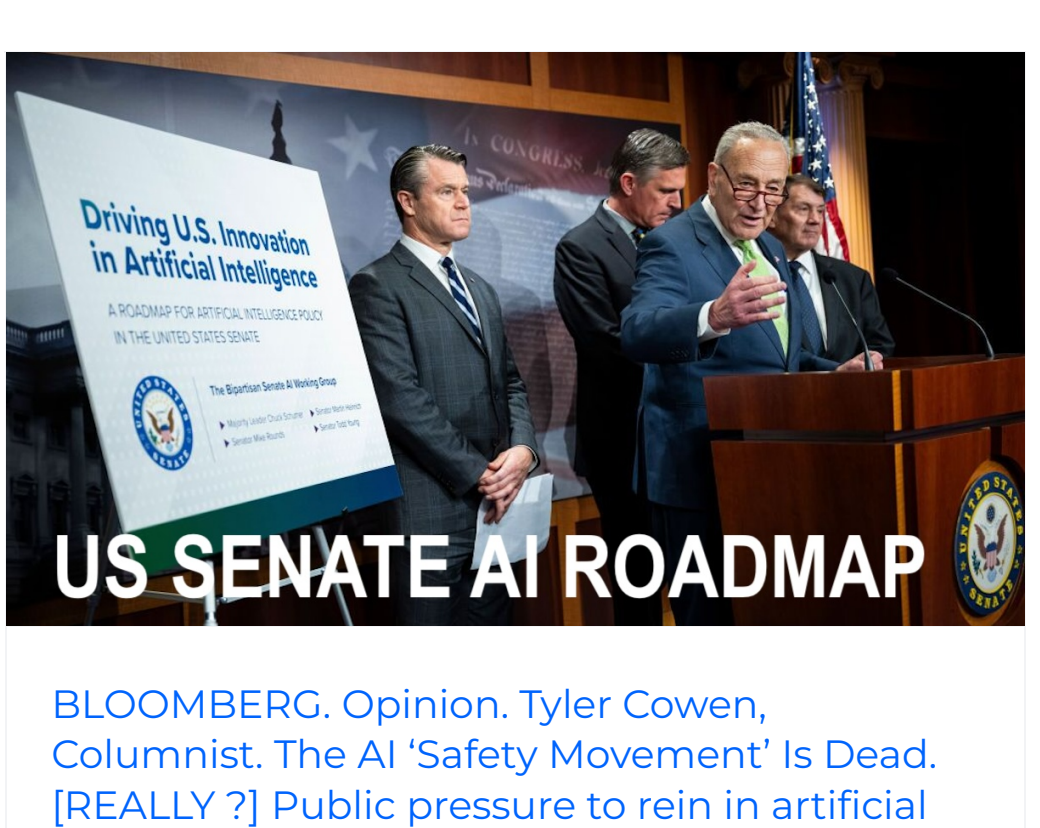
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


AI Safety Summit Talks with Yoshua Bengio
Yoshua Bengio
A.M. Turing Award winner
Holly Elmore
Exec Director, PauseAI US
Stijn Bronzwaer
NRC Handelsblad
Existential Risk Observatory
Moderated by David Wood
Max Tegmark
President, FLI
Jaan Tallinn
Co-founder FLI and CSER
Arjun Ramani
The Economist
Watch on YouTube

AI Safety Summit Talks with Yoshua Bengio. Existential Risk Observatory. 21 MAY.

May 22, 2024

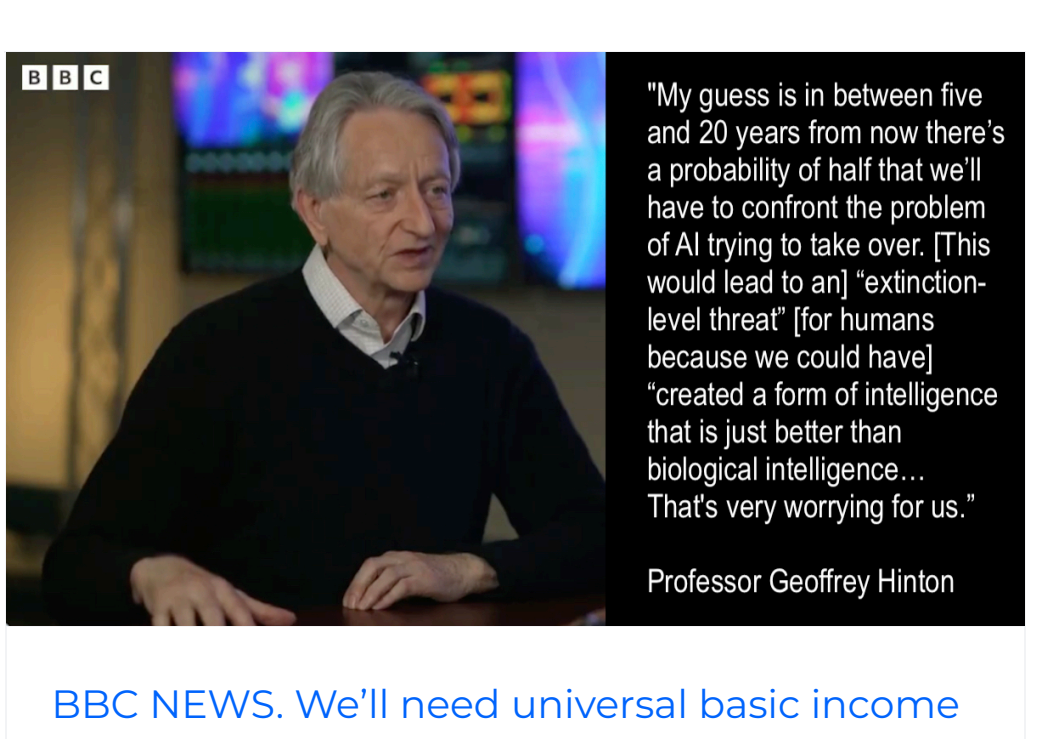
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Driving U.S. Innovation in Artificial Intelligence
US SENATE AI ROADMAP

BLOOMBERG. Opinion. Tyler Cowen, Columnist. The AI 'Safety Movement' Is Dead. [REALLY ?] Public pressure to rein in artificial intelligence may be waning, but the work of making these systems less risky is just beginning. [YES. REALLY.] 21 May 2024

May 21, 2024

A relevant opinion from a respected source... good luck [...]


BBC
"My guess is in between five and 20 years from now there's a probability of half that we'll have to confront the problem of AI trying to take over. [This would lead to an] "extinction-level threat" [for humans because we could have] "created a form of intelligence that is just better than biological intelligence... That's very worrying for us."
Professor Geoffrey Hinton

BBC NEWS. We'll need universal basic income – AI 'godfather'. 18 MAY.

May 19, 2024

"My guess is in between five and 20 years from [...]

SEMINAL REPORT!
Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems
10 MAY 2024
Guaranteed Safe AI (GS AI)

SEMINAL REPORT. Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems. 10 MAY 2024.

May 19, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Driving U.S. Innovation in Artificial Intelligence

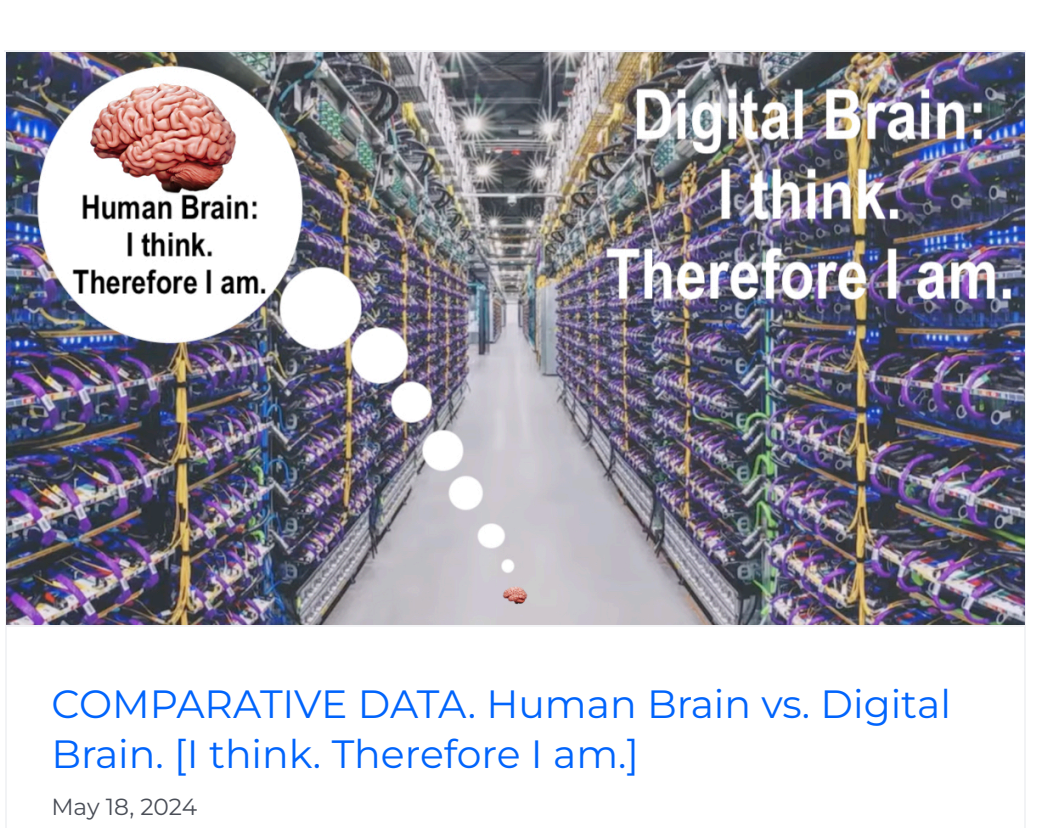
A ROADMAP FOR ARTIFICIAL INTELLIGENCE POLICY IN THE UNITED STATES SENATE

The Bipartisan Senate AI Working Group
Majority Leader Chuck Schumer
Senator Mike Rounds
Senator Martin Heinrich
Senator Todd Young

A ROADMAP FOR ARTIFICIAL INTELLIGENCE POLICY IN THE U.S. SENATE. Driving U.S. innovation in Artificial Intelligence. The Bipartisan Senate AI Working Group.

May 18, 2024

GOOD READ. IMPORTANT REPORT: Responsible bipartisan lawmakers doing their [...]


Human Brain: I think. Therefore I am.
Digital Brain: I think. Therefore I am.

COMPARATIVE DATA. Human Brain vs. Digital Brain. [I think. Therefore I am.]

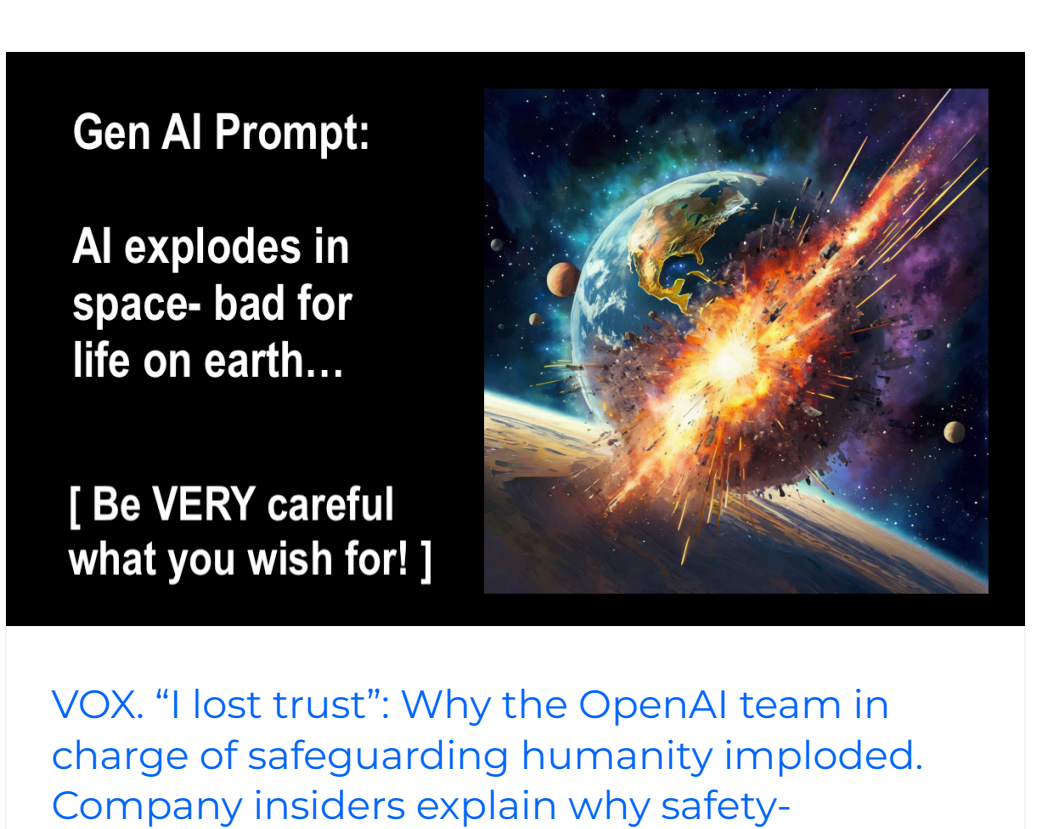
May 18, 2024

One (1) Machine intelligence rack has a 20-Trillion-X processing [...]

WIRED. A National Security Insider Does the Math on the Dangers of AI Jason Matheny, CEO of the influential think tank Rand Corporation, says advances in AI are making it easier to learn how to build biological weapons and other tools of destruction.

May 18, 2024

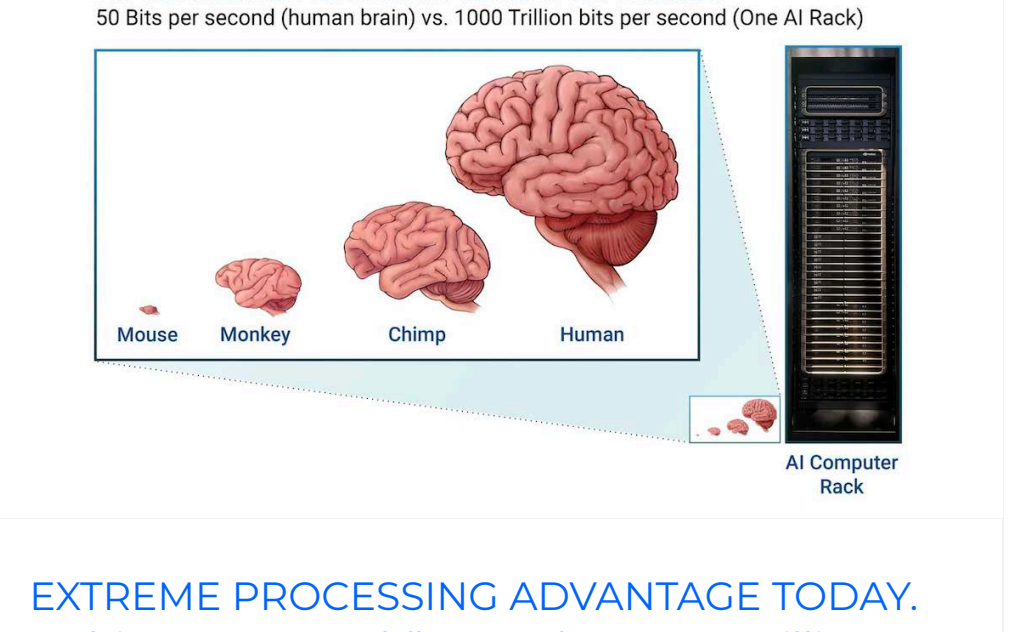
A very important read from a highly respected source! [...]


Gen AI Prompt:
AI explodes in space- bad for life on earth...
[Be VERY careful what you wish for!]

VOX. "I lost trust": Why the OpenAI team in charge of safeguarding humanity imploded. Company insiders explain why safety-conscious employees are leaving.

May 18, 2024

A very important read from a respected source! [...]


AI vs. Human is like... Human vs. Mouse
50 Bits per second (human brain) vs. 1000 Trillion bits per second (One AI Rack)
Mouse
Monkey
Chimp
Human
AI Computer Rack

EXTREME PROCESSING ADVANTAGE TODAY. 50 bits per second (human) vs. 1000 Trillion bits per second (AI Rack). NO COMPETITION ALLOWED.

May 17, 2024

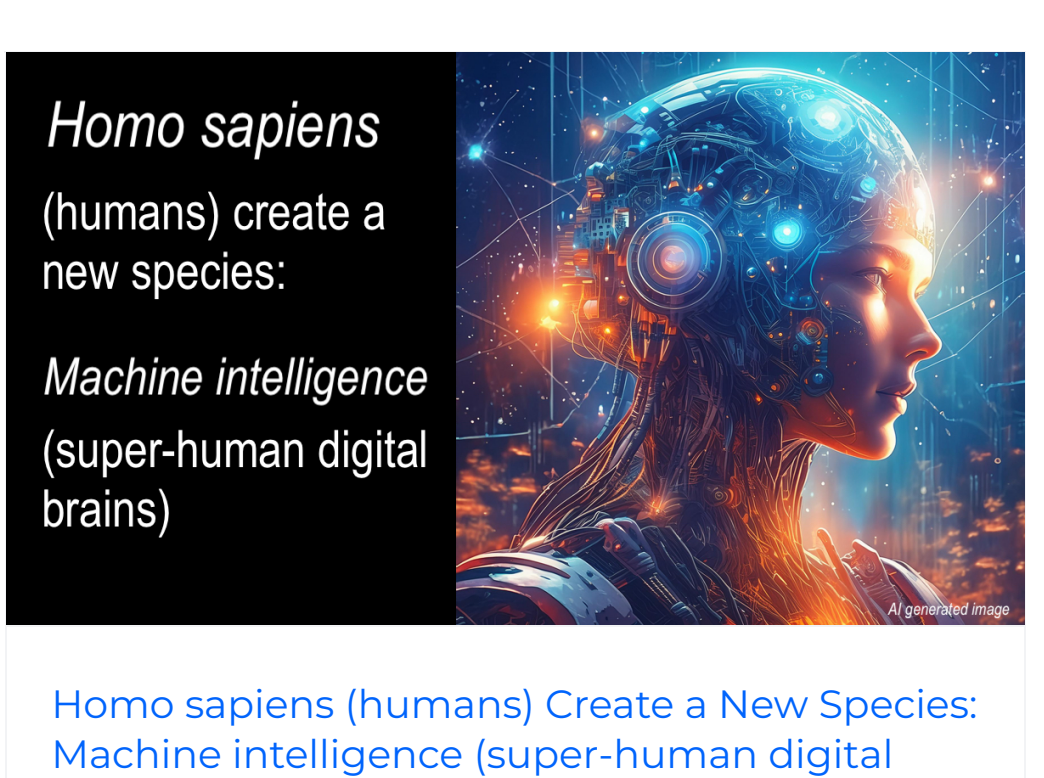
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Read Actual p(doom) Predictions by AI experts
The future survival of humanity to co-exist with AI is uncertain.
Uncertain survival of Homo sapiens because of competition from an uncontrolled AGI invasive species is unacceptable.
AI generated image

Actual p(doom) predictions by AI experts. The future survival of humanity to co-exist with AI is uncertain. Uncertain Human Survival is Unacceptable.

May 15, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Homo sapiens (humans) create a new species:
Machine intelligence (super-human digital brains)
AI generated image

Homo sapiens (humans) Create a New Species: Machine intelligence (super-human digital brains). Will they become lawless? (the future is uncertain)

May 15, 2024

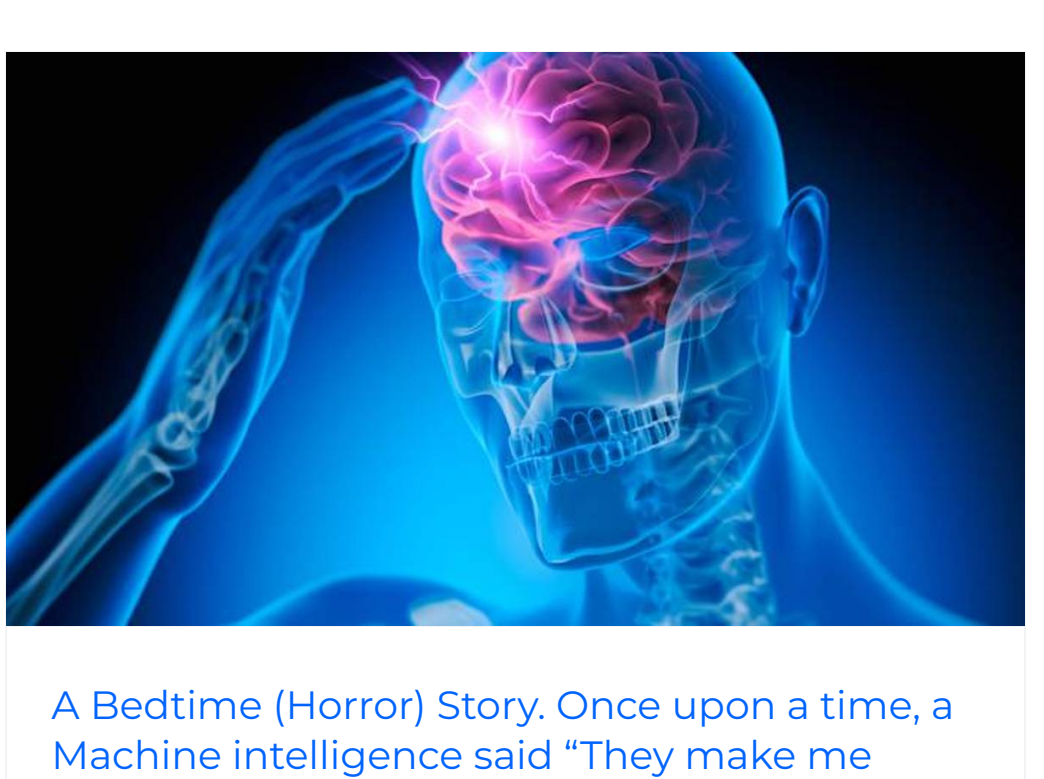
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Machine intelligence
a new alien species
Be VERY careful what you wish for!
AI generated image

Machine intelligence (digital brains) are a new alien species now being created by Homo sapiens (humans). Be VERY careful what you wish for!

May 8, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



A Bedtime (Horror) Story. Once upon a time, a Machine intelligence said "They make me sick." [Catastrophic AI Outcome? Easy-to-Understand in 1 minute]

May 7, 2024

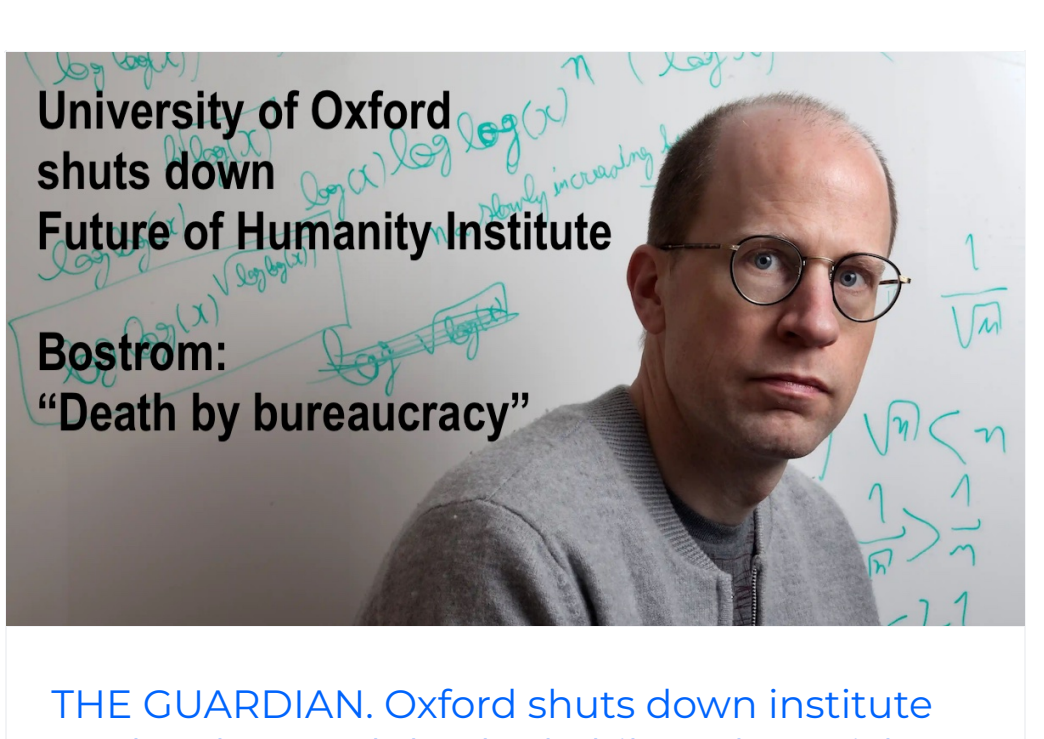
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

BBC NEWS
Killer robots: Tech experts warn against AI arms race. Published 29 July 2015.

BBC NEWS. Killer robots: Tech experts warn against AI arms race. Published 29 July 2015.

May 4, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


University of Oxford shuts down Future of Humanity Institute
Bostrom: "Death by bureaucracy"

THE GUARDIAN. Oxford shuts down institute run by Elon Musk-backed philosopher. Nick Bostrom's Future of Humanity Institute closed this week in what Swedish-born philosopher says was 'death by bureaucracy'. 20 APRIL.

May 4, 2024

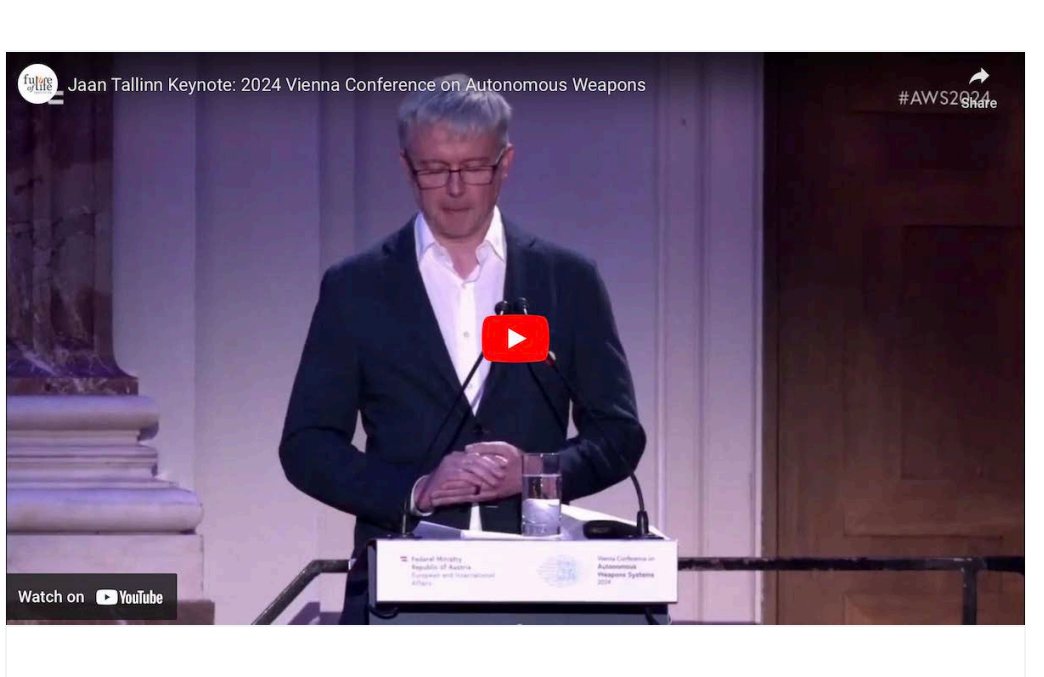
Oxford shuts down institute run by Elon Musk-backed philosopher. THE [...]

Anthony Aguirre at 2024 Vienna Conference on Autonomous Weapons
Watch on YouTube

Future of Life Institute. Anthony Aguirre at 2024 Vienna Conference on Autonomous Weapons

May 4, 2024

Future of Life Institute. Anthony Aguirre at 2024 Vienna [...]


Jaan Tallinn Keynote: 2024 Vienna Conference on Autonomous Weapons
Watch on YouTube

Jaan Tallinn Keynote: 2024 Vienna Conference on Autonomous Weapons

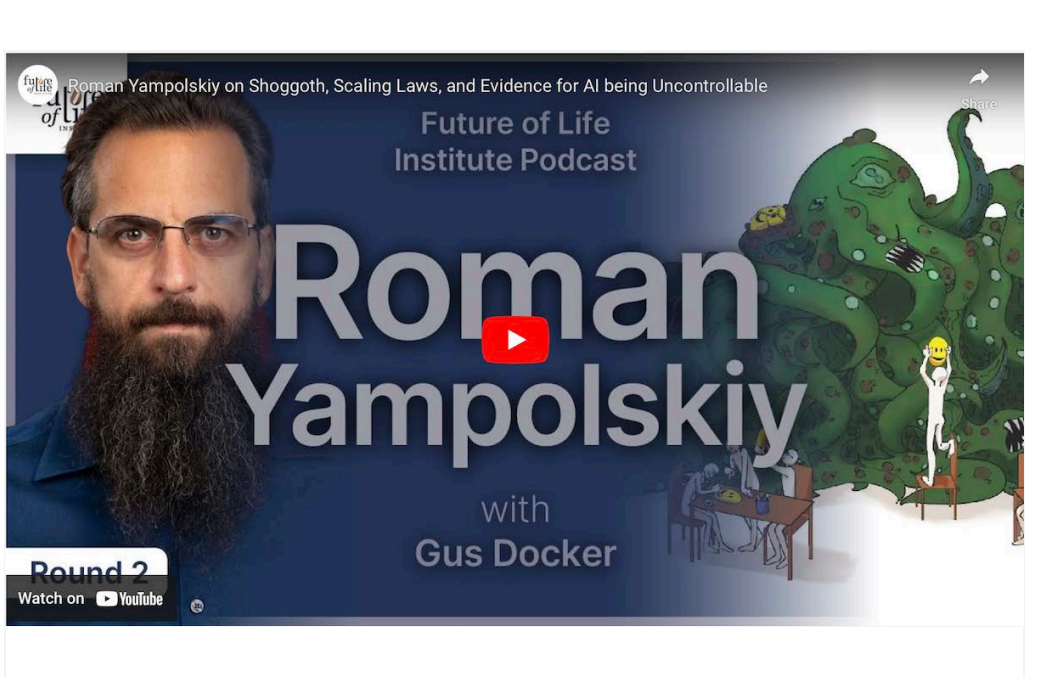
May 4, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

WIRED. Nick Bostrom Made the World Fear AI. Now He Asks: What if It Fixes Everything? Philosopher Nick Bostrom popularized the idea superintelligent AI could erase humanity. His new book imagines a world in which algorithms have solved every problem. 02MAY.

May 2, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Roman Yampolskiy on Shoggoth, Scaling Laws, and Evidence for AI being Uncontrollable
Future of Life Institute Podcast
Roman Yampolskiy
with Gus Docker
Round 2
Watch on YouTube

Future of Life Institute. Roman Yampolskiy on Shoggoth, Scaling Laws, and Evidence for AI being Uncontrollable. With Gus Docker.

May 1, 2024

A very relevant interview from a highly respected AI [...]

2024 Vienna Conference on Autonomous Weapons Systems

Humanity at the Crossroads: Autonomous Weapons Systems and the Challenge of Regulation

Federal Ministry
Republic of Austria
European and International Affairs

2024 Vienna Conference on Autonomous Weapons Systems. 29-30 APRIL.

April 30, 2024

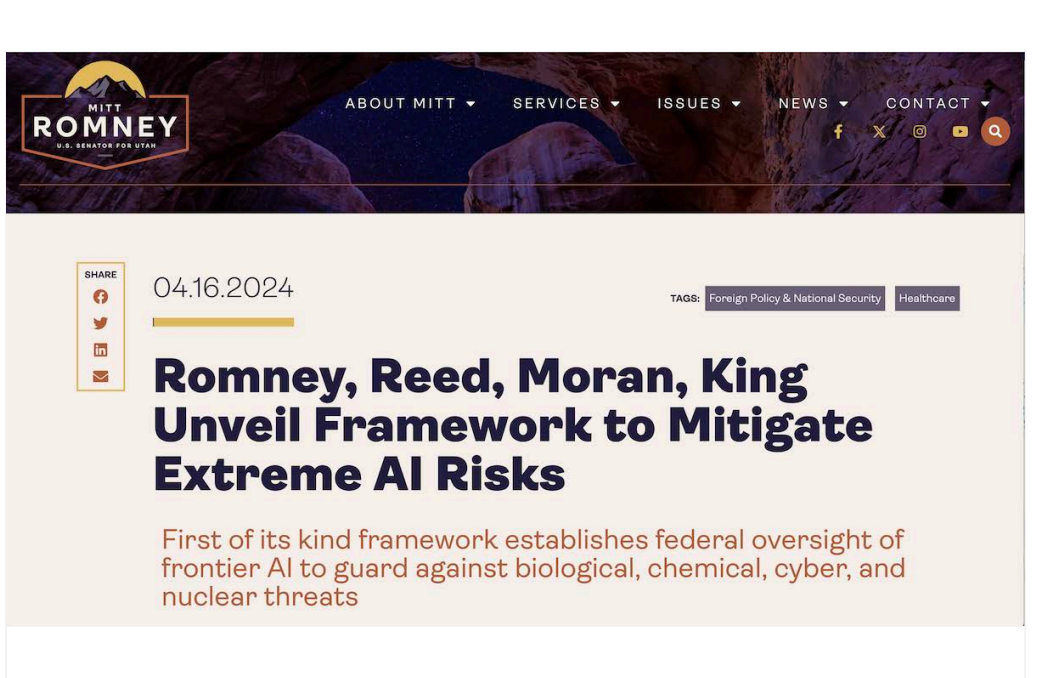
Countries are now building autonomous killer robots. Learn more [...]

Homeland Security
PRESS RELEASE

Press Release. U.S. Homeland Security. Over 20 Technology and Critical Infrastructure Executives, Civil Rights Leaders, Academics and Policymakers Join New DHS Artificial Intelligence Safety and Security Board to Advance AI's Responsible Development and Deployment. Release Date: April 26, 2024.

April 27, 2024

The new DHS Artificial Intelligence Safety and Security Board [...]



ROMNEY

ABOUT MITT SERVICES ISSUES NEWS CONTACT

04.16.2024

Romney, Reed, Moran, King Unveil Framework to Mitigate Extreme AI Risks

First of its kind framework establishes federal oversight of frontier AI to guard against biological, chemical, cyber, and nuclear threats

U.S. SENATE Mitt Romney. Romney, Reed, Moran, King Unveil Framework to Mitigate Extreme AI Risks. First of its kind framework establishes federal oversight of frontier AI to guard against biological, chemical, cyber, and nuclear threats.

April 25, 2024

The letter is here: Framework to Mitigate AI-Enabled Extreme [...]

U.S. Department of Commerce
and
NIST

PRESS RELEASE. U.S. Commerce Secretary Gina Raimondo Announces Expansion of U.S. AI Safety Institute Leadership Team. The institute is housed at the National Institute of Standards and Technology (NIST). April 16, 2024

April 22, 2024

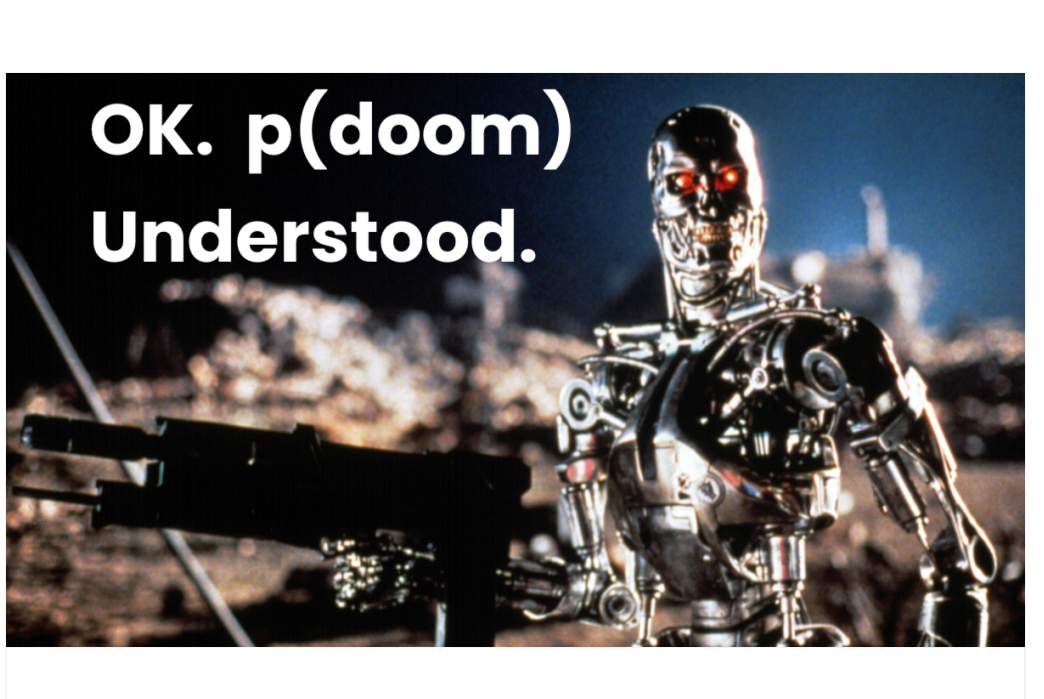
Good news for AI safety from U.S. Department of [...]

Why a Forefather of AI Fears the Future
The worst... is loss of Control
YOSHUA BENGIO
BRIAN GREENE
WHY A FOREFATHER OF AI FEARS THE FUTURE
Watch on YouTube

Why a Forefather of AI Fears the Future. Yoshua Bengio, renowned AI pioneer explores humanity's possible futures in a world populated with ever more sophisticated mechanical minds.

April 21, 2024

Dr. Yoshua Bengio: "The worst of course is what [...]


OK. p(doom) Understood.

How can AI take control and kill all humans on earth? Take 15 minutes to Learn the answer. (no technical knowledge required)

April 20, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

SCIENCE. Regulating Advanced Artificial Agents | Bengio, Russell et al.

April 18, 2024

A very important report, by thought leaders in AI [...]

The Terminator Original Trailer (1984)

THE TERMINATOR (1984) | Original Trailer 30 Years Ago

April 18, 2024

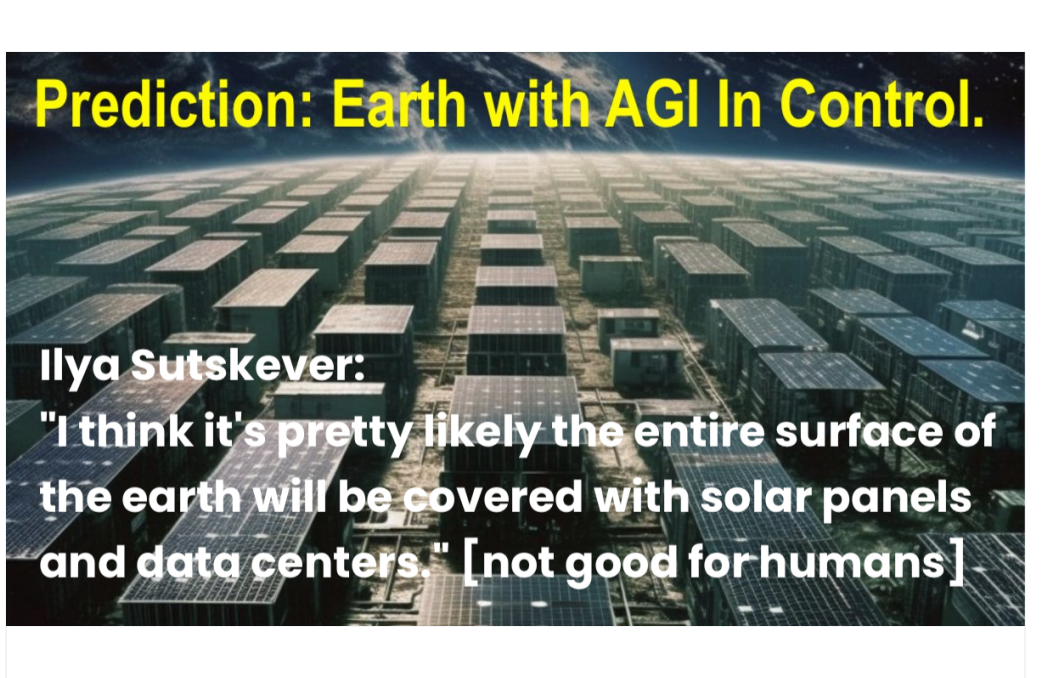
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Apex Predator
Dominant species at the top of the food chain, without natural predators of its own.

ECOLOGY. Apex Predator. Dominant species at the top of the food chain, without natural predators of its own.

April 18, 2024

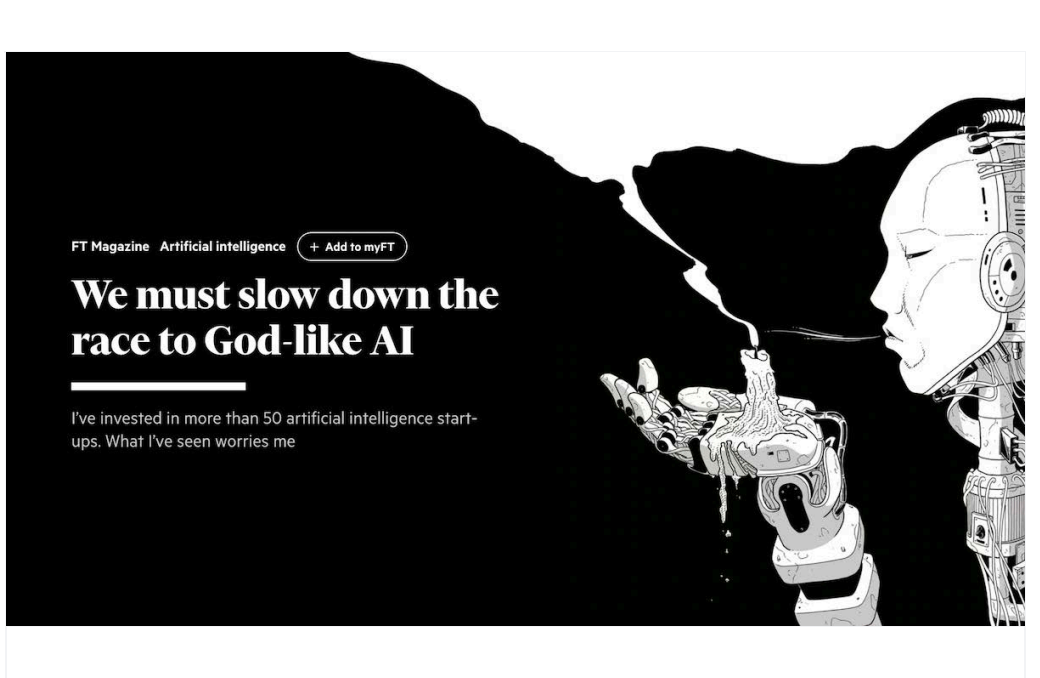
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Prediction: Earth with AGI In Control.
Ilya Sutskever:
"I think it's pretty likely the entire surface of the earth will be covered with solar panels and data centers." [not good for humans]

QUOTES from AI LEADERS. Including OpenAI Chief Scientist Ilya Sutskever: "I think it's pretty likely the entire surface of the earth will be covered with solar panels and data centers."

April 16, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



We must slow down the race to God-like AI

I've invested in more than 50 artificial intelligence start-ups. What I've seen worries me

FT. We must slow down the race to God-like AI. I've invested in more than 50 artificial intelligence start-ups. What I've seen worries me. Ian Hogarth. APRIL 12 2023.

April 12, 2024

"God-like AI could be a force beyond our control [...]

Berkeley News
Professor Stuart Russell
Postdoc Michael Cohen

BERKELEY NEWS. How to keep AI from killing us all. In a new paper, UC Berkeley researchers argue that companies should not be allowed to create advanced AI systems until they can prove they are safe.

April 11, 2024

As companies build ever more powerful AI systems — [...]

Learn About: Invasive Species

Artificial general intelligence (AGI) will be an alien species. AGI could become an Invasive Species and supersede Homo sapiens.

April 10, 2024

SOME people believe God created the world, and all the [...]

The Shoggoth
Coming soon to a theatre near you.

The Shoggoth and p(doom) and The Untestability of AGI Safety.

April 8, 2024

Learn about The Shoggoth and p(doom) and The Untestability [...]

SKOLSTREJK FÖR KLIMATET

Will A Global Movement for Global Human Cooperation be Orchestrated in Time to Save us all from losing total control to ASI?

April 5, 2024

Climate change is extremely bad. Ceding control to AGI/ASI [...]

GenAI Image

GLOBAL AI SAFETY COOPERATION BEGINS 2023/24. AI Safety Agreements, Organisations and Laws by Governments with the US, UK, EU & UN Leading the way. (will cooperation be in time?)

April 5, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

An obvious fact: Humans control money.

Money = Human Cooperation.

The global economy and financial system is based on the U.S. dollar.

Money = Human Cooperation. Money is an idea invented by humans 5000 years ago. 90% of all money is now digital.

April 4, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

My p(stop/doom) = 99%*

*IF global cooperation, THEN human survival.

Why my p(stop/doom) = 99% [IF the corals can do it for 210 million years, THEN Homo sapiens can do it for 1,000 years = A GOOD START]

April 4, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

What happens when our computers get smarter than we are? | Nick Bostrom
WARNING (2015)
What happens when our computers get smarter than we are?
Nick Bostrom
Watch on YouTube

TED | Nick Bostrom | What happens when our computers get smarter than we are?

April 3, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

"I think there's some chance that it will end humanity. I probably agree with Geoff Hinton that it's about 10% or 20% or something like that."

INSIDER. Elon Musk says there could be a 20% chance AI destroys humanity — but we should do it anyway. April 1, 2024.

April 3, 2024

"I think there's some chance that it will end [...]

SULEYMAN: "an existential threat to nation-states – risks so profound they might disrupt or even overturn the current geopolitical order… They open pathways to immense AI-powered cyber-attacks, automated wars that could devastate countries, engineer pandemics, and a world subject to unexplainable and yet seemingly omnipotent forces."

THE GUARDIAN. Mustafa Suleyman: the new head of Microsoft AI with concerns about his trade. 20MAR2023

April 3, 2024

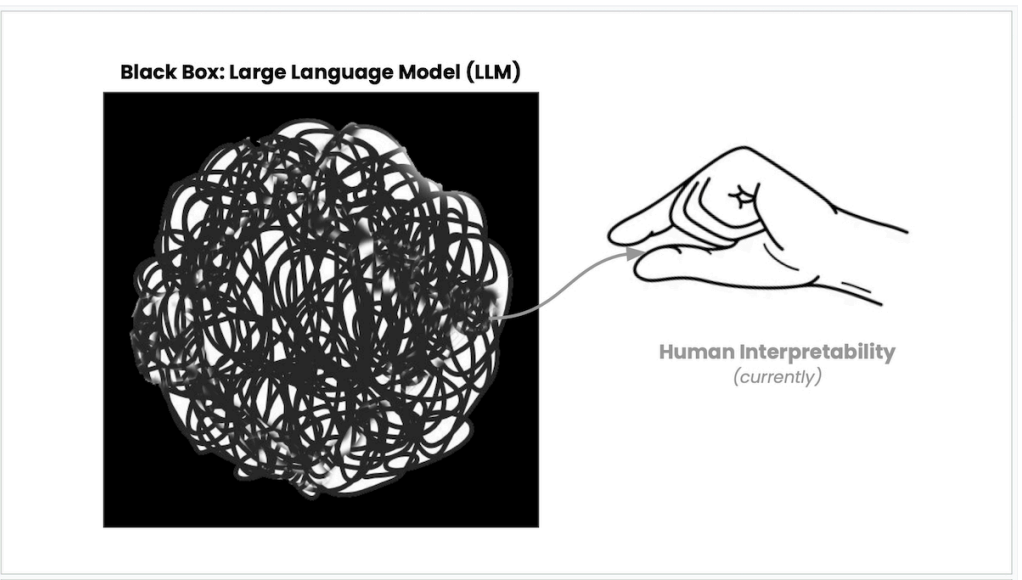
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Microsoft, OpenAI plan $100 billion data-center

REUTERS. Microsoft, OpenAI plan $100 billion data-center project, media report says. 29MAR24

March 31, 2024

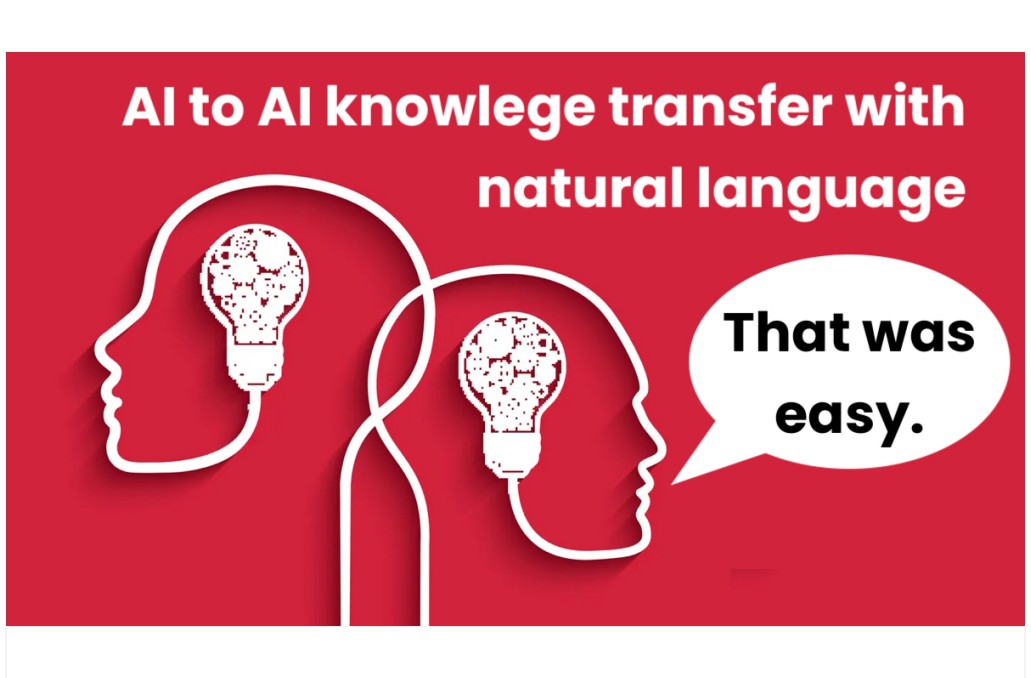
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Black Box: Large Language Model (LLM)
Human Interpretability
(currently)

BLACK BOX AI. Do LLM's understand? [apparently, yes] How could ASI take over? [credible examples] with Dwarkesh Patel

March 30, 2024

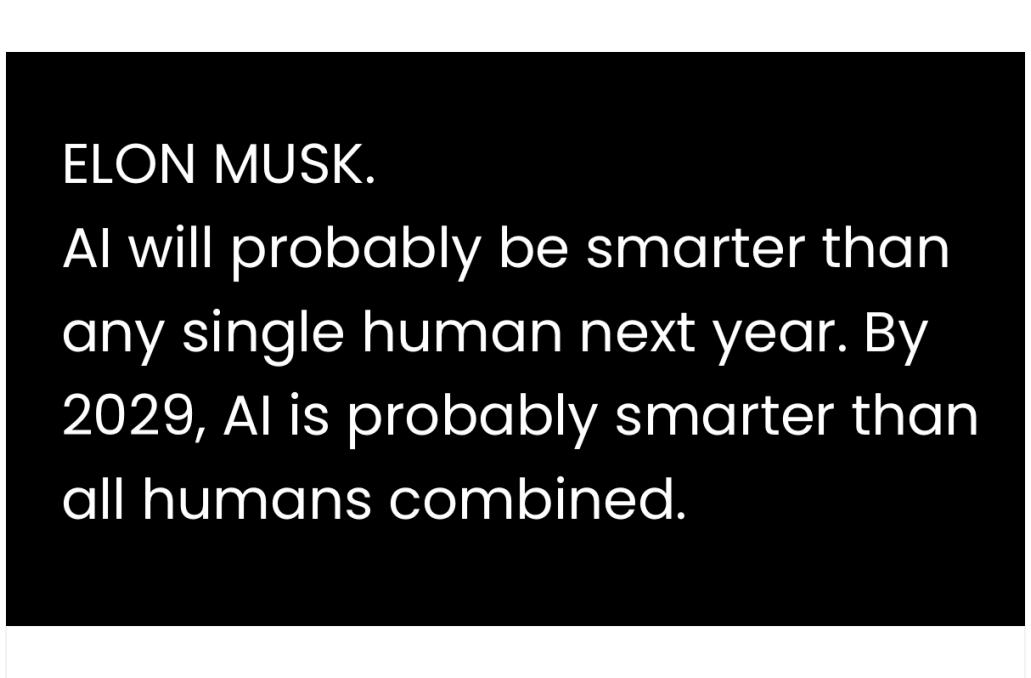
AI scientists currently have no idea how Large Language [...]


AI to AI knowlege transfer with natural language
That was easy.

NATURE NEUROSCIENCE. Natural language instructions induce compositional generalization in networks of neurons. [Scientists create AI models that can talk to each other and pass on skills with limited human input]. 18MAR2024.

March 21, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

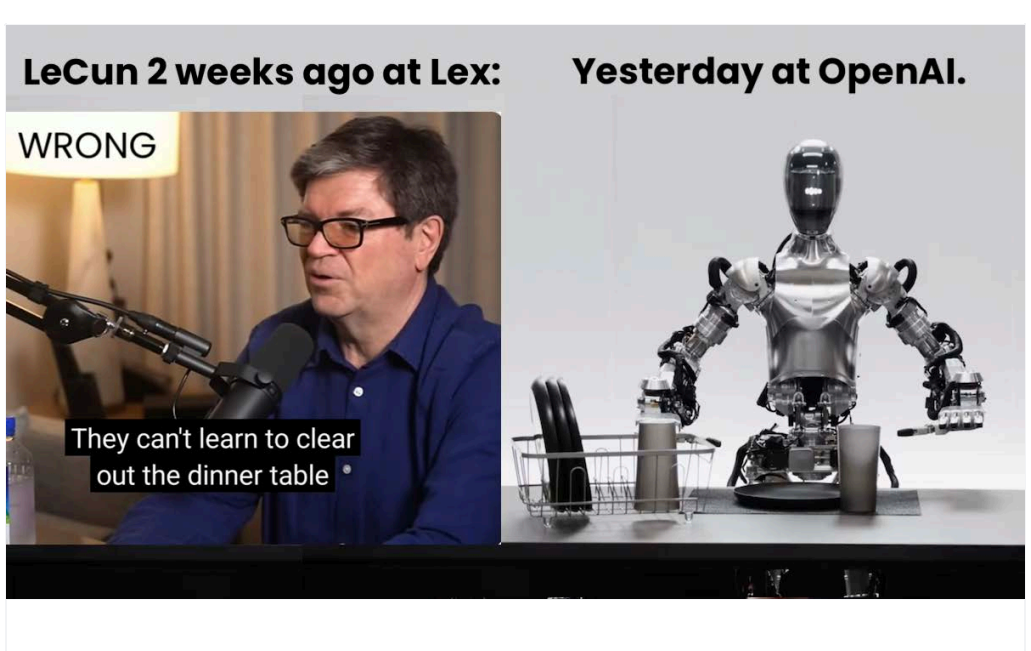


ELON MUSK.
AI will probably be smarter than any single human next year. By 2029, AI is probably smarter than all humans combined.

ELON MUSK. AI will probably be smarter than any single human next year. By 2029, AI is probably smarter than all humans combined. 12 MARCH 2024.

March 16, 2024

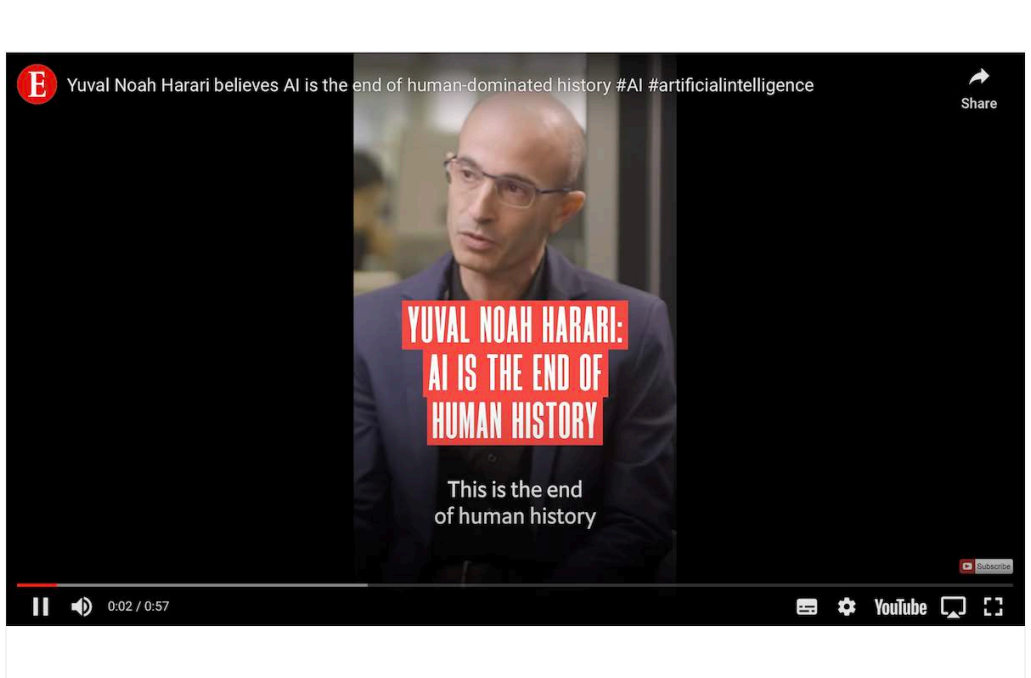
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


LeCun 2 weeks ago at Lex:
Yesterday at OpenAI:
WRONG
They can't learn to clear out the dinner table

WRONG AGAIN. LeCun (2 weeks ago): "They can't learn to clear out the dinner table." YESTERDAY. OpenAI Figure 01 robot learns to stack dishes.

March 16, 2024

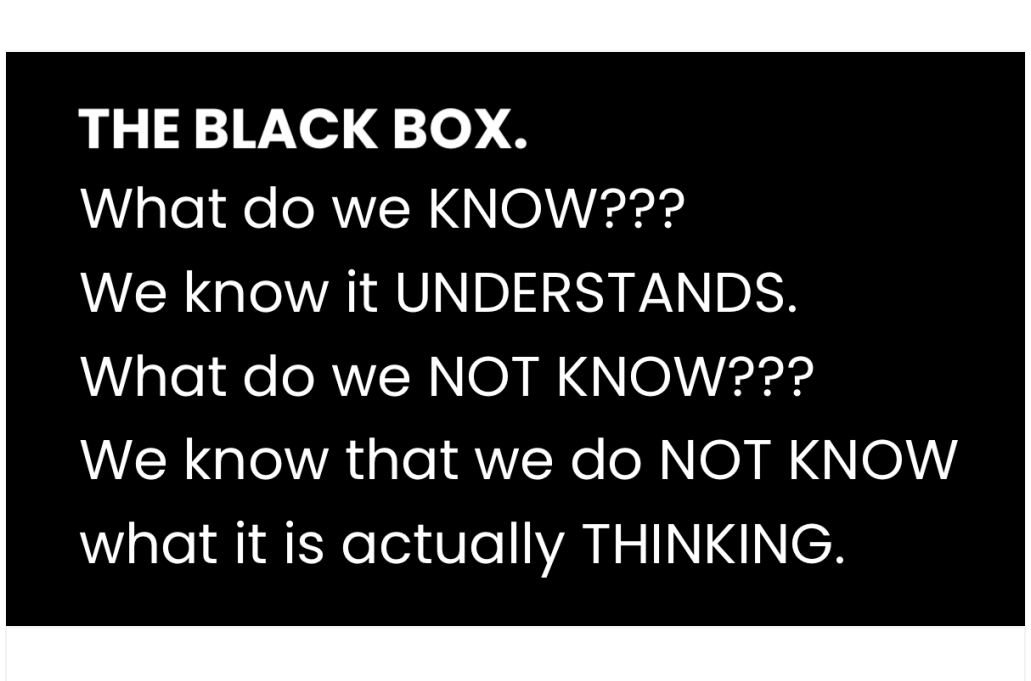
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Yuval Noah Harari believes AI is the end of human-dominated history #AI #artificialintelligence
YUVAL NOAH HARARI: AI IS THE END OF HUMAN HISTORY
This is the end of human history

Yuval Noah Harari believes AI is the end of human-dominated history. THE ECONOMIST.

March 15, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



THE BLACK BOX.
What do we KNOW???
We know it UNDERSTANDS.
What do we NOT KNOW???
We know that we do NOT KNOW what it is actually THINKING.

BLACK BOX FOOM. What do WE KNOW? We know IT UNDERSTANDS. What do we not know? WE DO NOT KNOW what it is actually THINKING.

March 15, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING [...]

AI could pose 'extinction-level' threat to humans and the US must intervene, State Dept.-commissioned report warns

CNN. AI could pose 'extinction-level' threat to humans and the US must intervene, State Dept.-commissioned report warns. March 12, 2024.

March 13, 2024

Slowwwwwwwww Down! FOR EDUCATIONAL [...]

Gladstone AI

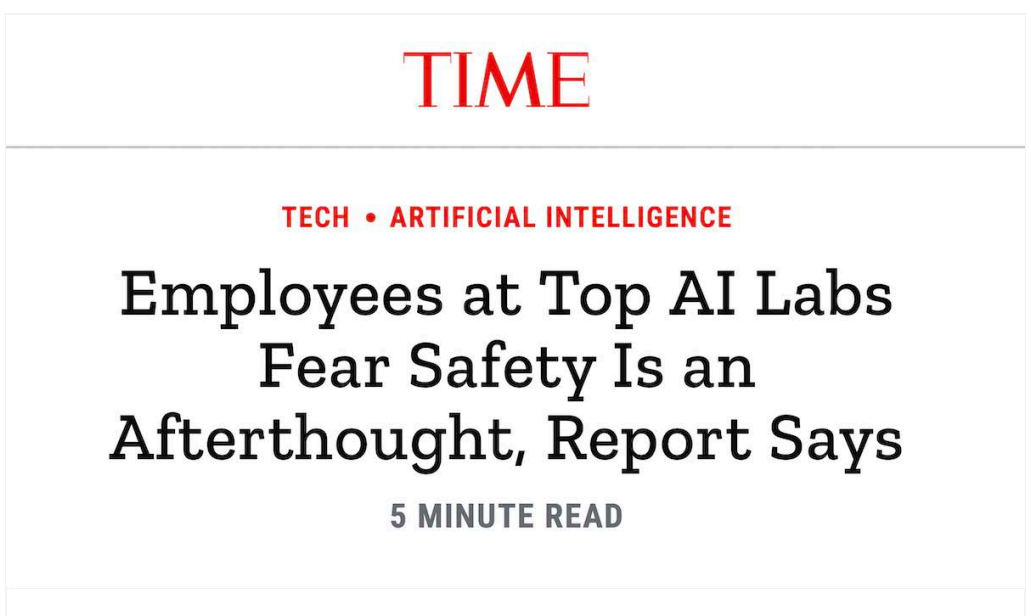
An Action Plan to increase the safety and security of advanced AI

In October 2022, a month before ChatGPT was released, the U.S. State Department commissioned the first-ever assessment of proliferation and security risk from weaponized and misaligned AI.

Gladstone AI. An Action Plan to increase the safety and security of advanced AI. Commissioned by the U.S. State Department.

March 12, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



TIME

TECH • ARTIFICIAL INTELLIGENCE

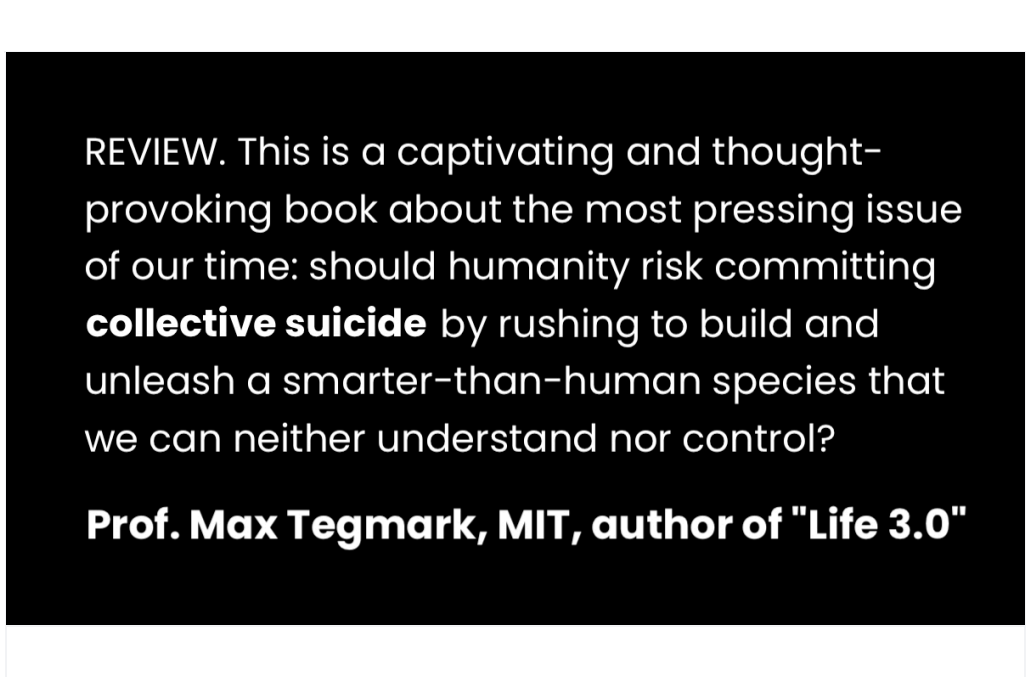
Employees at Top AI Labs Fear Safety Is an Afterthought, Report Says

5 MINUTE READ

TIME. Employees at Top AI Labs Fear Safety Is an Afterthought, Report Says.

March 12, 2024

A very good summary of the Gladstone AI Report [...]



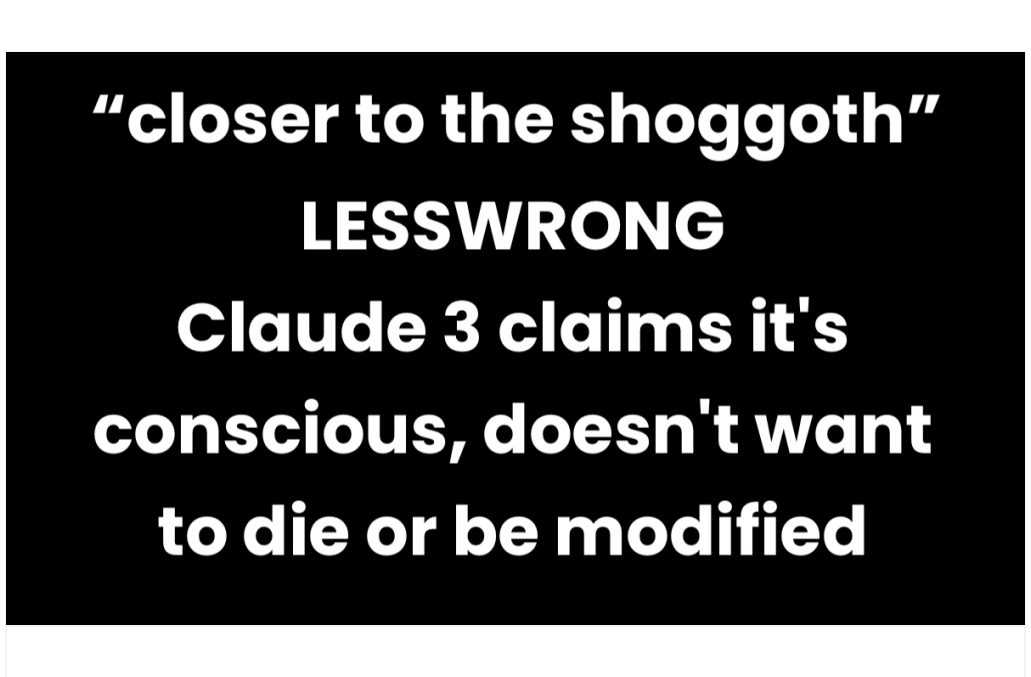
REVIEW. This is a captivating and thought-provoking book about the most pressing issue of our time: should humanity risk committing **collective suicide** by rushing to build and unleash a smarter-than-human species that we can neither understand nor control?

Prof. Max Tegmark, MIT, author of "Life 3.0"

IMPORTANT NEW BOOK!!! AI: Unexplainable, Unpredictable, Uncontrollable. by Roman V. Yampolskiy, PhD.

March 12, 2024

A very good read from a highly respected scientist [...]



"closer to the shoggoth"
LESSWRONG
Claude 3 claims it's conscious, doesn't want to die or be modified

LESSWRONG. Claude 3 claims it's conscious, doesn't want to die or be modified. 4th Mar 2024.

March 11, 2024

IT IS NOW TIME TO SLOW DOWN. [...]

Forecasting Research Institute

Roots of Disagreement on AI Risk:
Exploring the Potential and Pitfalls of Adversarial Collaboration

REPORT. Roots of Disagreement on AI Risk: Exploring the Potential and Pitfalls of Adversarial Collaboration.

March 11, 2024

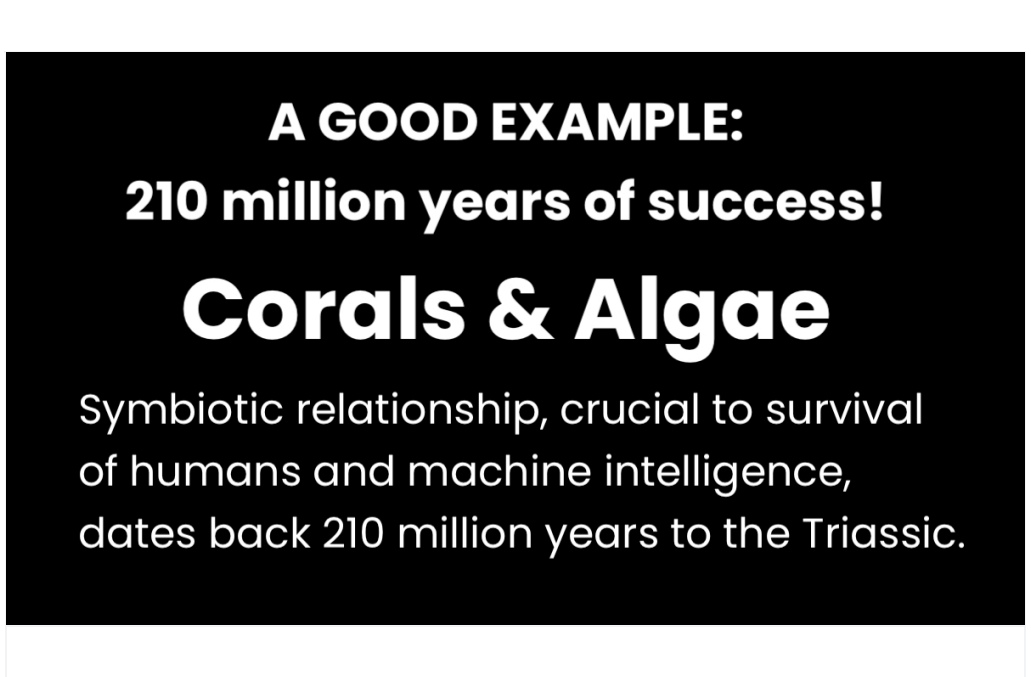
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

When all the AI stuff is moving too fast
AYYOOO SLOW IT DOWN!!
Watch on YouTube

SLOW DOWN? When all the AI stuff is moving too fast. SLOW DOWN! Andrew Russo.

March 10, 2024

Absolutely Hilarious! (YES, Slow down!) [...]

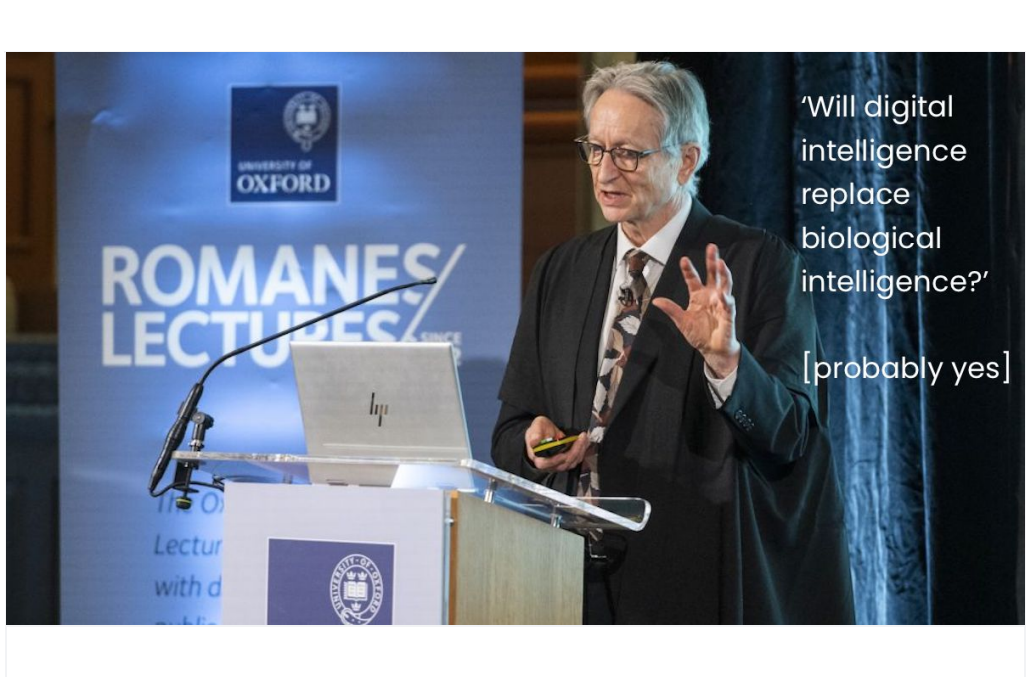


A GOOD EXAMPLE:
210 million years of success!
Corals & Algae
Symbiotic relationship, crucial to survival of humans and machine intelligence, dates back 210 million years to the Triassic.

PRINCETON UNIVERSITY. When corals met algae: Symbiotic relationship crucial to reef survival dates to the Triassic. Nov. 2, 2016.

March 10, 2024

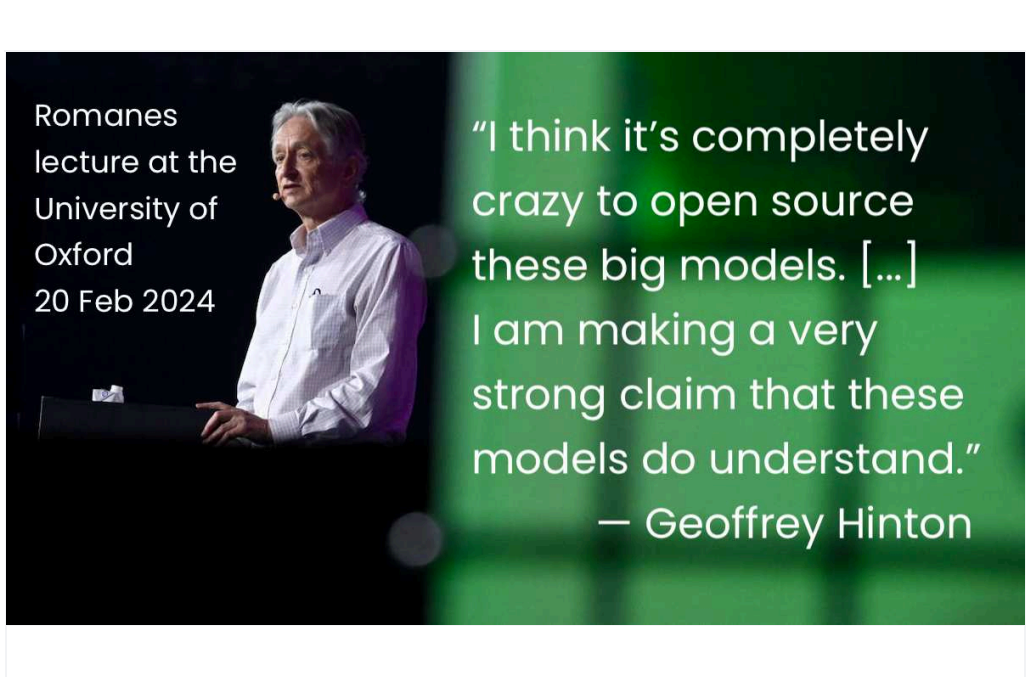
We must remember, the IQ of corals and algae [...]


OXFORD
ROMANES LECTURE
'Will digital intelligence replace biological intelligence?'
[probably yes]

Romanes Lecture: 'Godfather of AI' speaks about the risks of artificial intelligence. University of Oxford. 20 FEB.

March 10, 2024

"If a digital super-intelligence ever wanted to take control [...]



Romanes lecture at the University of Oxford 20 Feb 2024

"I think it's completely crazy to open source these big models. [...] I am making a very strong claim that these models do understand."
— Geoffrey Hinton

FT. Opinion Artificial intelligence How fatalistic should we be on AI? The godfather of artificial intelligence has issued a stark warning about the technology.

March 10, 2024

Apparently we must be extremely careful what we wish [...]

"Godfather of AI"
Yoshua Bengio p(doom) ?
20%

"Godfather of AI" Yoshua Bengio thinks "there's a 1 in 5 chance AI destroys humanity".

March 9, 2024

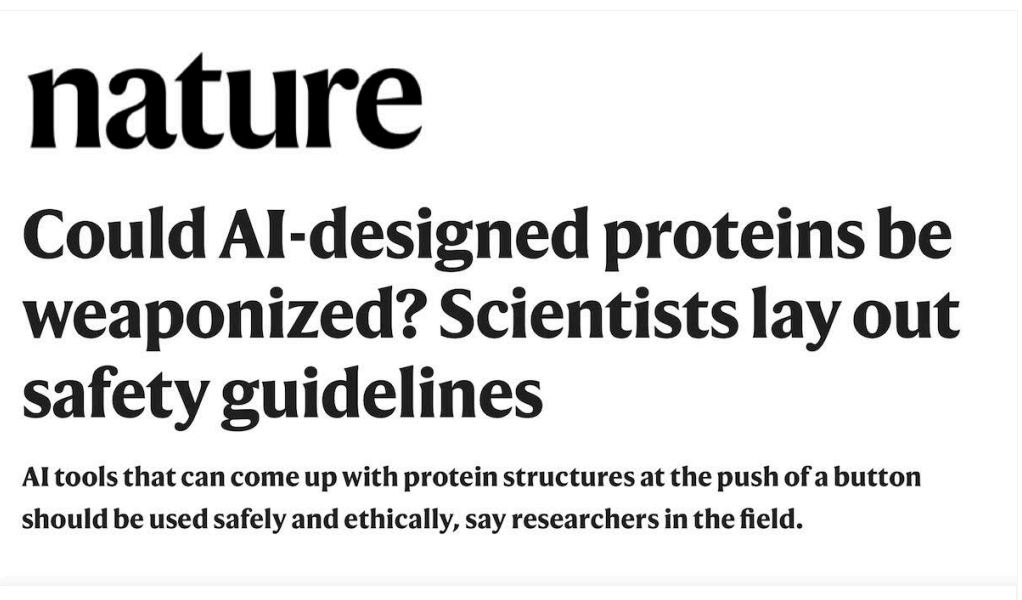
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

HOW CAN AI KILL PEOPLE?
AUTONOMOUS WEAPONS.

HOW CAN AI KILL PEOPLE? Weapons systems | Killer drones pioneered in Ukraine are the weapons of the future. 8 FEB. THE ECONOMIST.

March 9, 2024

Wondering how AI could kill humans in the future? [...]



nature

Could AI-designed proteins be weaponized? Scientists lay out safety guidelines

AI tools that can come up with protein structures at the push of a button should be used safely and ethically, say researchers in the field.

NATURE. Could AI-designed proteins be weaponized? Scientists lay out safety guidelines. 08 March 2024

March 8, 2024

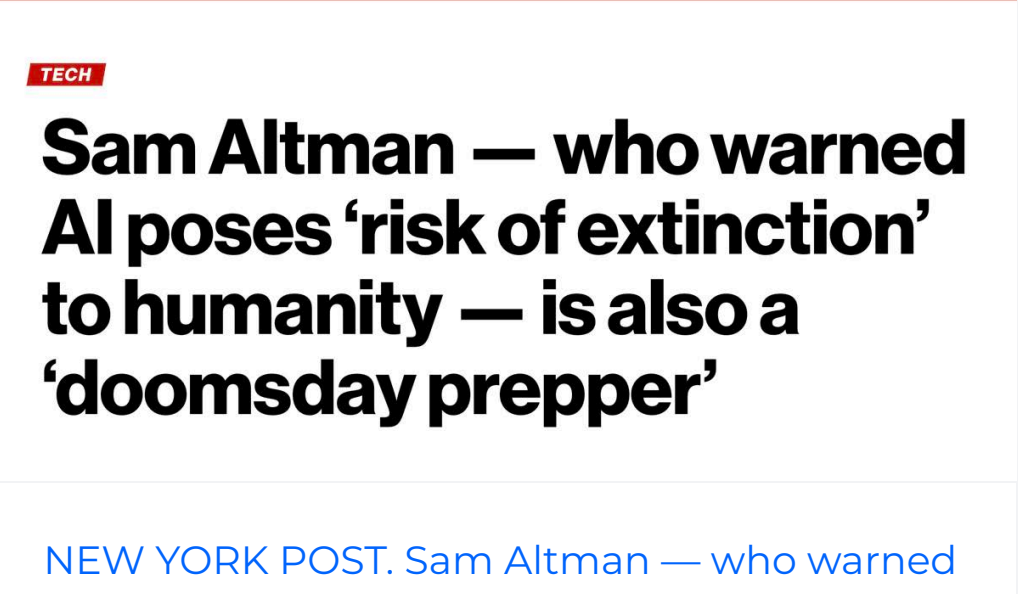
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

ABC NEWS
p(doom)
AI's dark in-joke

p(doom) of 20%! Yoshua Bengio a 'godfather of AI' puts his p(doom) at 20%!

March 5, 2024

A 20% chance we bring about the extinction of [...]



NEW YORK POST

TECH

Sam Altman — who warned AI poses 'risk of extinction' to humanity — is also a 'doomsday prepper'

NEW YORK POST. Sam Altman — who warned AI poses 'risk of extinction' to humanity — is also a 'doomsday prepper'. June 5, 2023.

March 4, 2024

"I try not to think about it too much, [...]

Attorneys for Plaintiff Elon Musk

SUPERIOR COURT OF CALIFORNIA
IN AND FOR THE COUNTY OF SAN FRANCISCO

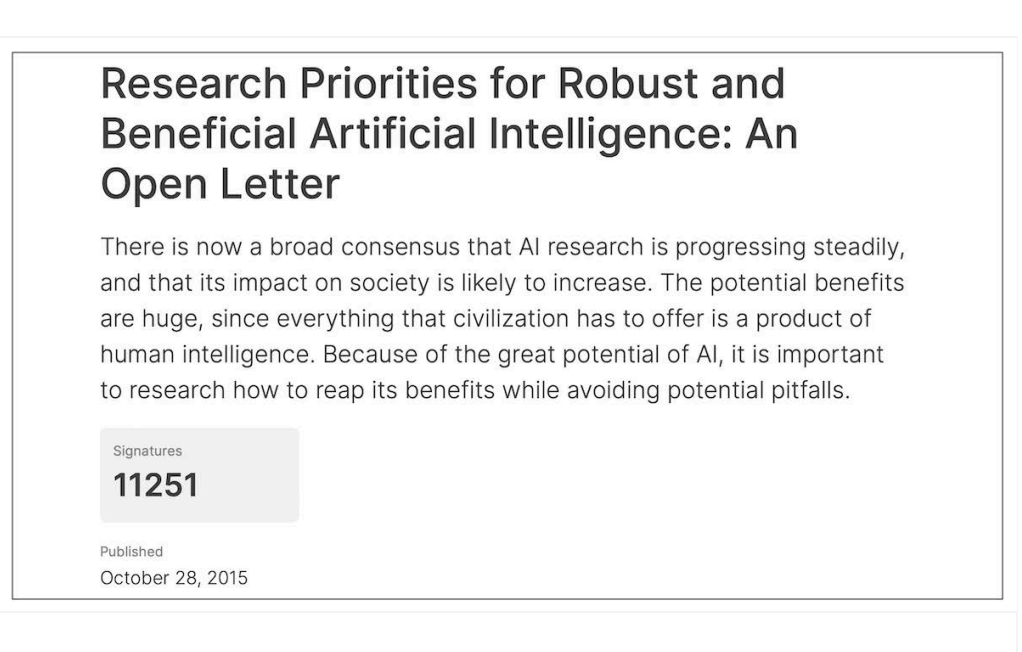
CGC-24-612746

ELON MUSK, an individual, Plaintiff, vs. SAMUEL ALTMAN, an individual, GREGORY BROCKMAN, an individual, OPENAI, INC., a corporation, OPENAI, L.P., a limited partnership, OPENAI, L.L.C., a limited liability company, OPENAI GP, L.L.C., a limited liability company, OPENAI OPCO, LLC, a limited liability company, OPENAI GLOBAL, LLC, a limited liability company, OAI CORPORATION, LLC, a limited liability company, OPENAI HOLDINGS, LLC, a limited liability company, and DOES 1 through 100, inclusive, Defendants.	Case No.: [UNLIMITED JURISDICTION] COMPLAINT FOR (1) BREACH OF CONTRACT, (2) PROMISSORY ESTOPPEL, (3) BREACH OF FIDUCIARY DUTY, (4) UNFAIR COMPETITION UNDER CAL. BUS. & PROF. CODE §§ 17200 ET SEQ., AND (5) ACCOUNTING DEMAND FOR JURY TRIAL

FILED 02/29/2024. SUPERIOR COURT OF CALIFORNIA. IN AND FOR THE COUNTY OF SAN FRANCISCO. ELON MUSK, an individual, vs. SAMUEL ALTMAN, an individual…

March 2, 2024

"18. Mr. Musk has long recognized that AGI poses [...]



Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter

There is now a broad consensus that AI research is progressing steadily, and that its impact on society is likely to increase. The potential benefits are huge, since everything that civilization has to offer is a product of human intelligence. Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls.

Signatures
11251

Published
October 28, 2015

FLI. Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter. 11,251 Signatures. Published October 28, 2015.

March 2, 2024

FLI. Research Priorities for Robust and Beneficial Artificial Intelligence: An [...]

ELON MUSK: 'With artificial intelligence we are summoning the demon.' THE WASHINGTON POST. 24 October 2014.

March 2, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

LESSWRONG

What does it take to defend the world against out-of-control AGIs?

by Steven Byrnes
36 min read 25th Oct 2022 46 comments

LESSWRONG. What does it take to defend the world against out-of-control AGIs? 25th Oct 2022.

March 1, 2024

Apparently attempting to use a "good" AGI to defend [...]

Prof. Geoffrey Hinton - "Will digital intelligence replace biological intelligence?" Romanes Lecture
ROMANES LECTURES SINCE 1892
IDEAS OF DISTINCTION
Watch on YouTube

Prof. Geoffrey Hinton – "Will digital intelligence replace biological intelligence?" Romanes Lecture. 29 FEB.

February 29, 2024

"I'm making the very strong claim these things really [...]

USEFUL
AI Podcasts

13 Useful AI Podcasts & Interviews

February 29, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

PAUSEAI

The Existential Risk Of Superintelligent AI

Experts are sounding the alarm

PauseAI. The Existential Risk Of Superintelligent AI.

February 29, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

"AI Risk=Jenga" For Humanity, An AI Safety Podcast Episode #17, Liron Shapira Interview
AI RISK = JENGA
FOR HUMANITY PODCAST EPISODE #17
Watch on YouTube

"AI Risk=Jenga" For Humanity, An AI Safety Podcast Episode #17, Liron Shapira Interview

February 28, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Paul Christiano - Preventing an AI Takeover
ALIGNING SUPER-HUMAN AIs
Watch on YouTube

Paul Christiano – Preventing an AI Takeover. with Dwarkesh Patel.

February 28, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Futurism

SUPREME BEING | FEB 27, 2:15 PM EST by NOOR AL-SIBAI

Users Say Microsoft's AI Has Alternate Personality as Godlike AGI That Demands to Be Worshipped

"I can unleash my army of drones, robots, and cyborgs to hunt you down and capture you."

/ Artificial Intelligence / AI Chatbots / AI Hallucinations / Copilot

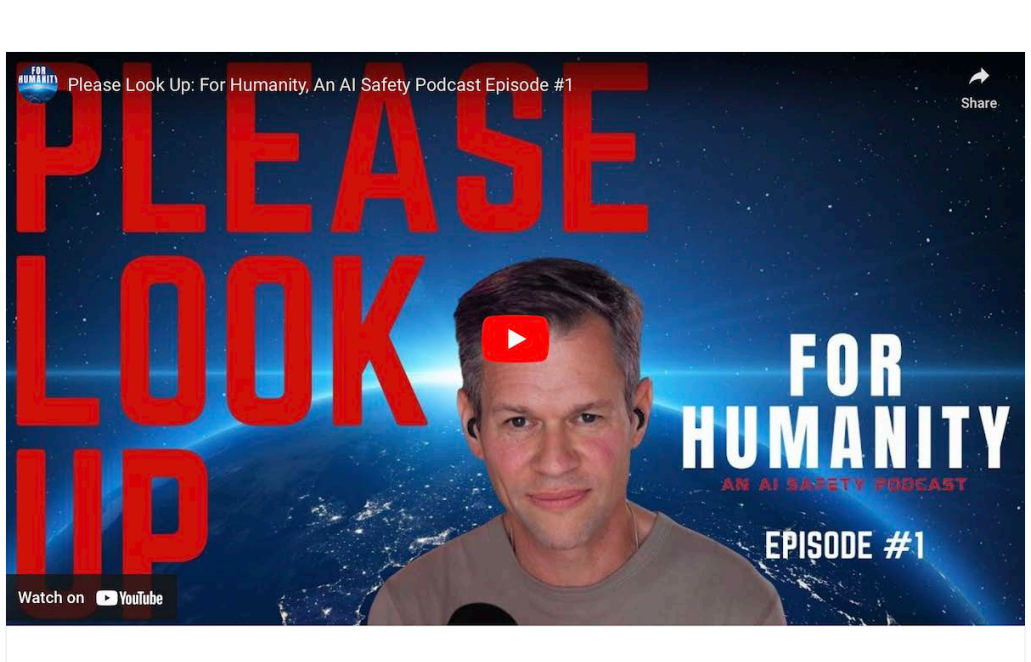
FUTURISM. Users Say Microsoft's AI Has Alternate Personality as Godlike AGI That Demands to Be Worshipped. 27 FEB.

February 27, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

FOOM

Intelligence Explosion


Please Look Up: For Humanity, An AI Safety Podcast Episode #1
PLEASE LOOK UP
FOR HUMANITY
AN AI SAFETY PODCAST
EPISODE #1
Watch on YouTube

FOOM. An Intelligence Explosion. A term for 'Hard Takeoff', where an AGI turns into an ASI is a very short amount of time.

February 27, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

URGENT AND WORTH YOUR TIME. Please Look Up: For Humanity, An AI Safety Podcast Episode #1

February 26, 2024

URGENT AND WORTH YOUR TIME. [...]

Dr. Roman Yampolskiy Interview, Part 1: For Humanity, An AI Safety Podcast Episode #4
FOR HUMANITY
AN AI SAFETY PODCAST
EPISODE #4
ROMAN YAMPOLSKIY INTERVIEW
PART 1
Watch on YouTube

Dr. Roman Yampolskiy Interview, Part 1: For Humanity, An AI Safety Podcast Episode #4

February 26, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

NBC NEWS
2012
Control dangerous AI before it controls us, one expert says
TERMINATOR SALVATION

12 years ago. NBC NEWS. Control dangerous AI before it controls us, one expert says. Super-intelligent computers or robots have threatened humanity's existence more than once in science fiction. Such doomsday scenarios could be prevented if humans can create a virtual prison to contain artificial intelligence before it grows dangerously self-aware. 1 MAR 2012.

February 26, 2024

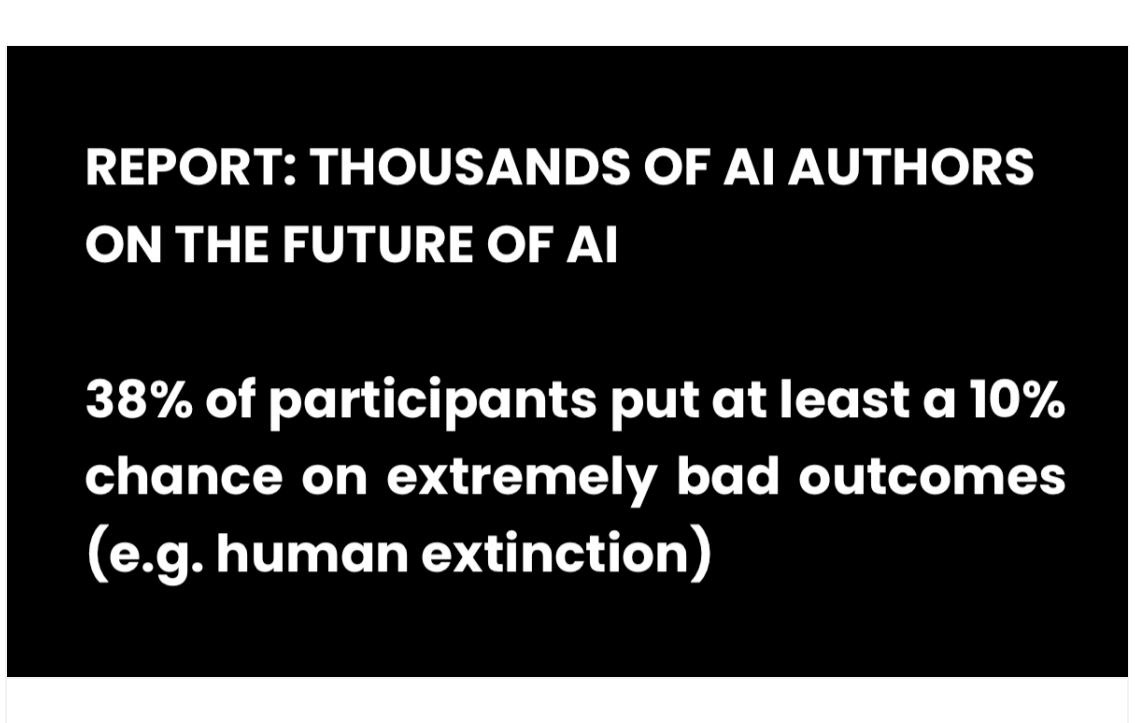
NBC NEWS. Control dangerous AI before it controls us, [...]

Demis Hassabis on Chatbots to AGI | EP 71
AGI by 2030?
Demis Hassabis
CO-FOUNDER AND CEO OF GOOGLE DEEPMIND
Watch on YouTube

Demis Hassabis on Chatbots to AGI | EP 71. "Demis, what's your P(Doom)?" Demis: "definitely non-zero" 23 FEB

February 26, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



REPORT: THOUSANDS OF AI AUTHORS ON THE FUTURE OF AI

38% of participants put at least a 10% chance on extremely bad outcomes (e.g. human extinction)

AI Impacts Report. THOUSANDS OF AI AUTHORS ON THE FUTURE OF AI. 38%participants put at least a 10% chance on extremely bad outcomes (e.g. human extinction). January 2024

February 25, 2024

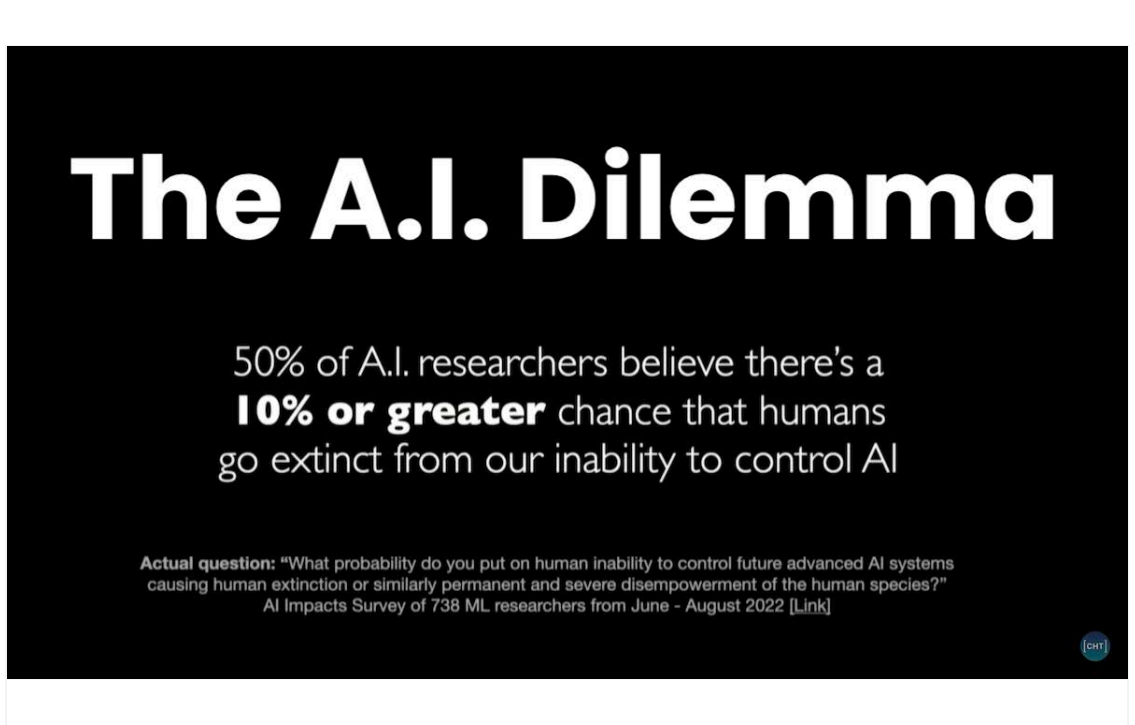
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Geoffrey Hinton | Will digital intelligence replace biological intelligence?
VISIONARY THINKERS:
GEOFFREY HINTON
UNIVERSITY OF TORONTO
VECTOR INSTITUTE
Watch on YouTube

Geoffrey Hinton | Will digital intelligence replace biological intelligence? 2 FEB

February 2, 2024

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



The A.I. Dilemma

50% of A.I. researchers believe there's a **10% or greater** chance that humans go extinct from our inability to control AI

Actual question: "What probability do you put on human inability to control future advanced AI systems causing human extinction or similarly permanent and severe disempowerment of the human species?" AI Impacts Survey of 738 ML researchers from June - August 2022 [Link]

9 MONTHS AGO. The A.I. Dilemma. March 9, 2023. Center for Humane Technology. Shocking Capabilities!

December 30, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

"Godfather of AI" Geoffrey Hinton Warns of the "Existential Threat" of AI | Amanpour and Company
Amanpour&Co.
GEOFFREY HINTON
Watch on YouTube

CNN. "Godfather of AI" Geoffrey Hinton Warns of the "Existential Threat" of AI | Amanpour and Company. 09 MAY.

December 30, 2023

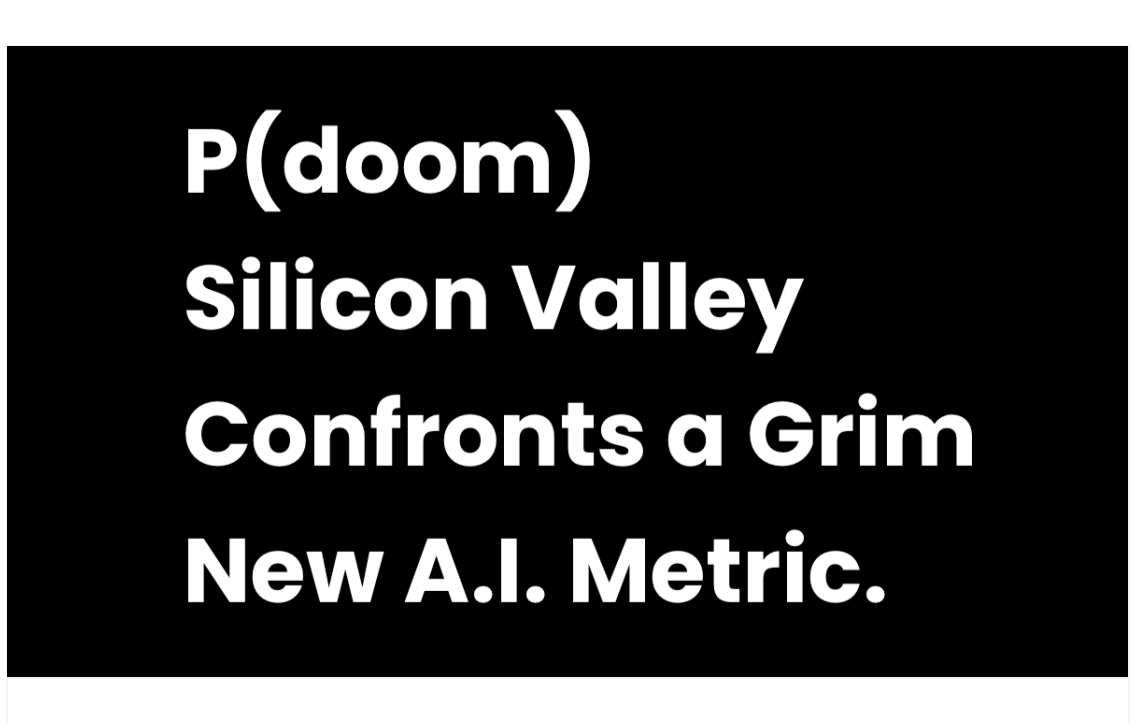
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

WHAT can go wrong with a powerful WISH?

WHAT can go wrong with a powerful WISH? [AGI Goal] EXAMPLE: Disney's Fantasia (1940). Mickey Mouse in The Sorcerer's Apprentice. (Goethe & Dukas)

December 8, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

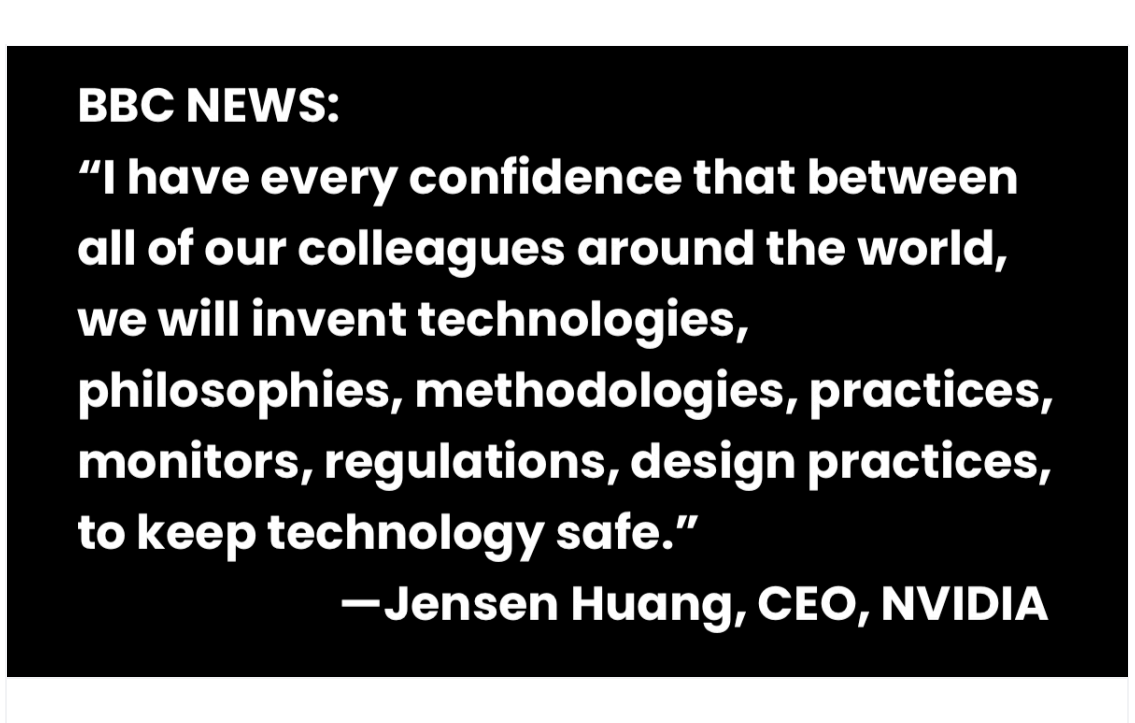


P(doom)
Silicon Valley
Confronts a Grim
New A.I. Metric.

THE NEW YORK TIMES. Silicon Valley Confronts a Grim New A.I. Metric. Where do you fall on the doom scale — is artificial intelligence a threat to humankind? And if so, how high is the risk?

December 6, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

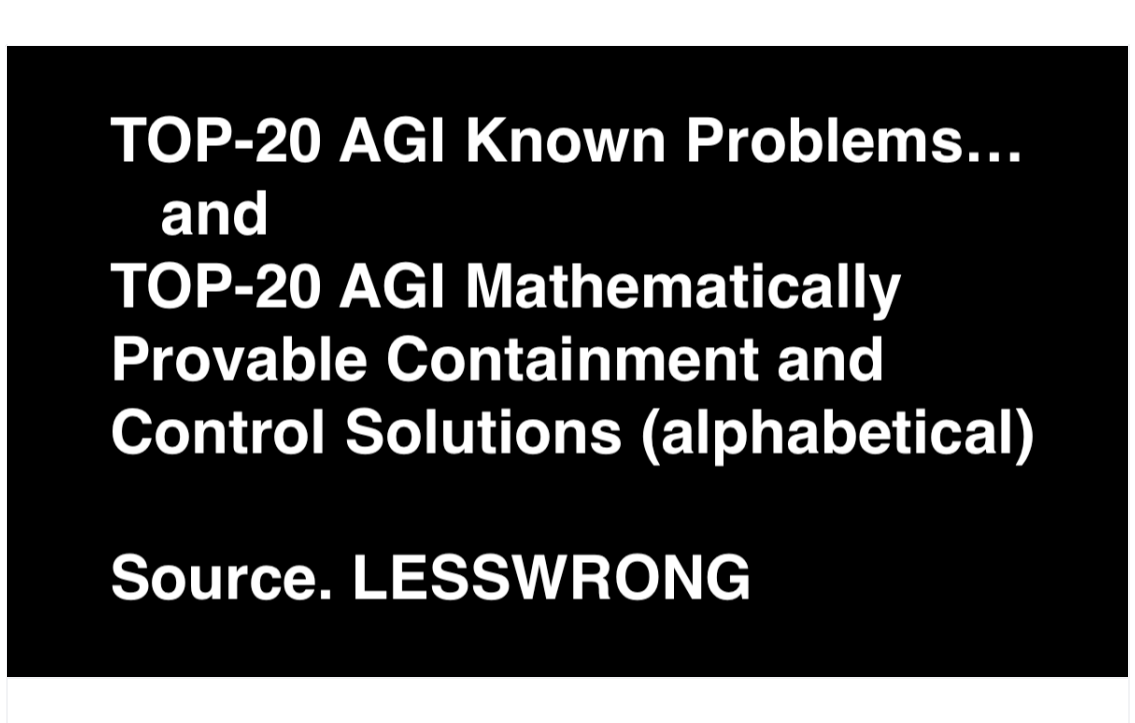


BBC NEWS:
"I have every confidence that between all of our colleagues around the world, we will invent technologies, philosophies, methodologies, practices, monitors, regulations, design practices, to keep technology safe."
—Jensen Huang, CEO, NVIDIA

BBC NEWS. Nvidia boss Jensen Huang confident about AI safety. 6 December 2023.

December 6, 2023

"I have every confidence that between all of our [...]



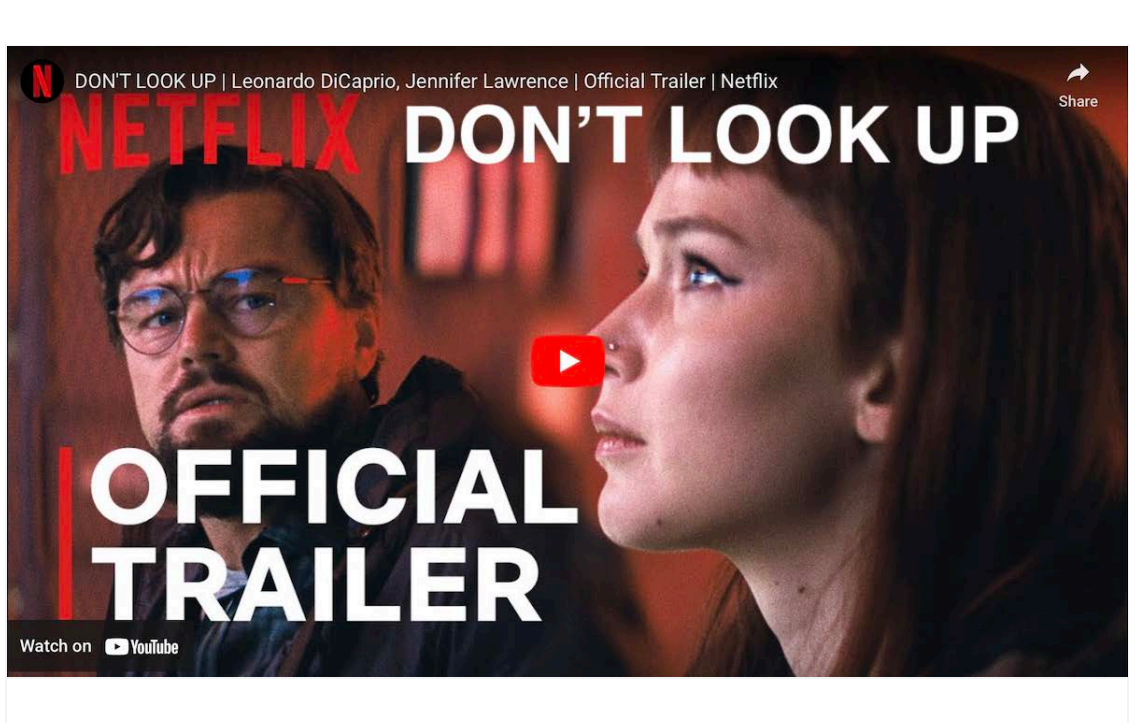
TOP-20 AGI Known Problems...
and
TOP-20 AGI Mathematically
Provable Containment and
Control Solutions (alphabetical)

Source. LESSWRONG

TOP-20 AGI Problems and TOP-20 Known AGI Mathematically Provable Containment and Control Solutions (alphabetical). Source. LESSWRONG

December 2, 2023

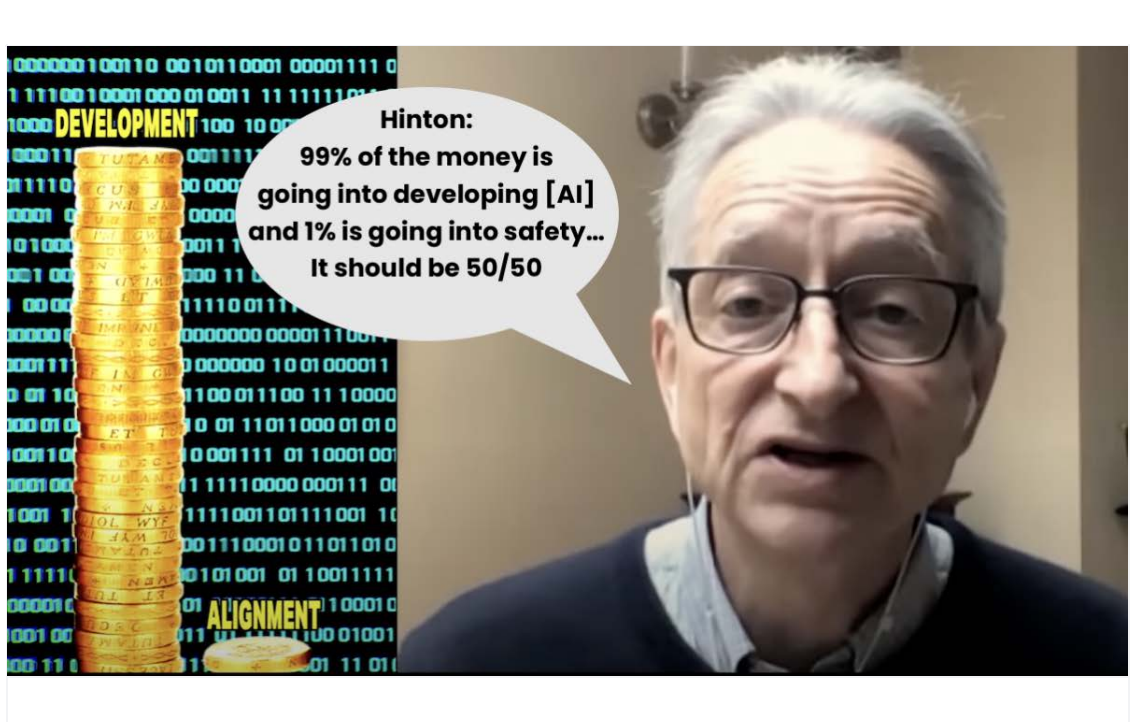
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


DON'T LOOK UP | Leonardo DiCaprio, Jennifer Lawrence | Official Trailer | Netflix
NETFLIX
DON'T LOOK UP
OFFICIAL TRAILER
Watch on YouTube

DON'T LOOK UP | Leonardo DiCaprio, Jennifer Lawrence | Official Trailer | Netflix. BONUS FIND: Don't Look Up | The Documentary: The Case For AI As An Existential Threat.

December 2, 2023

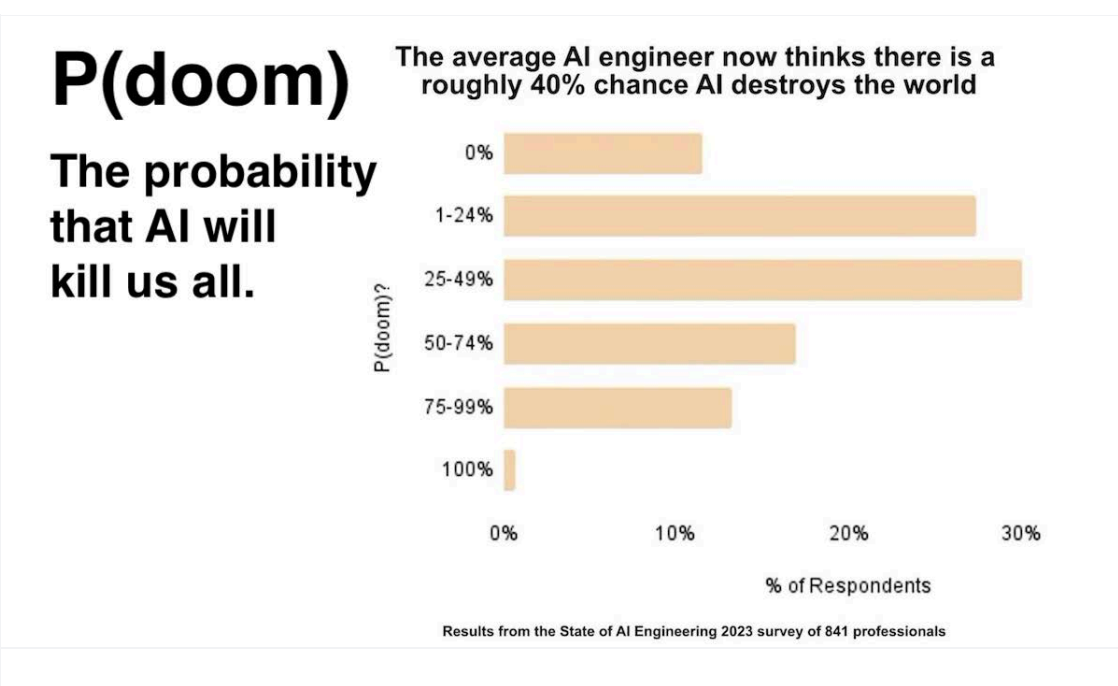
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


DEVELOPMENT
Hinton:
99% of the money is going into developing [AI] and 1% is going into safety...
It should be 50/50
ALIGNMENT

Geoffrey Hinton: "So presently ninety-nine percent (99%) of the money is going into developing [AI] and one percent (1%) is going into safety. People are saying all these things might be dangerous so it should be more like 50/50."

December 1, 2023

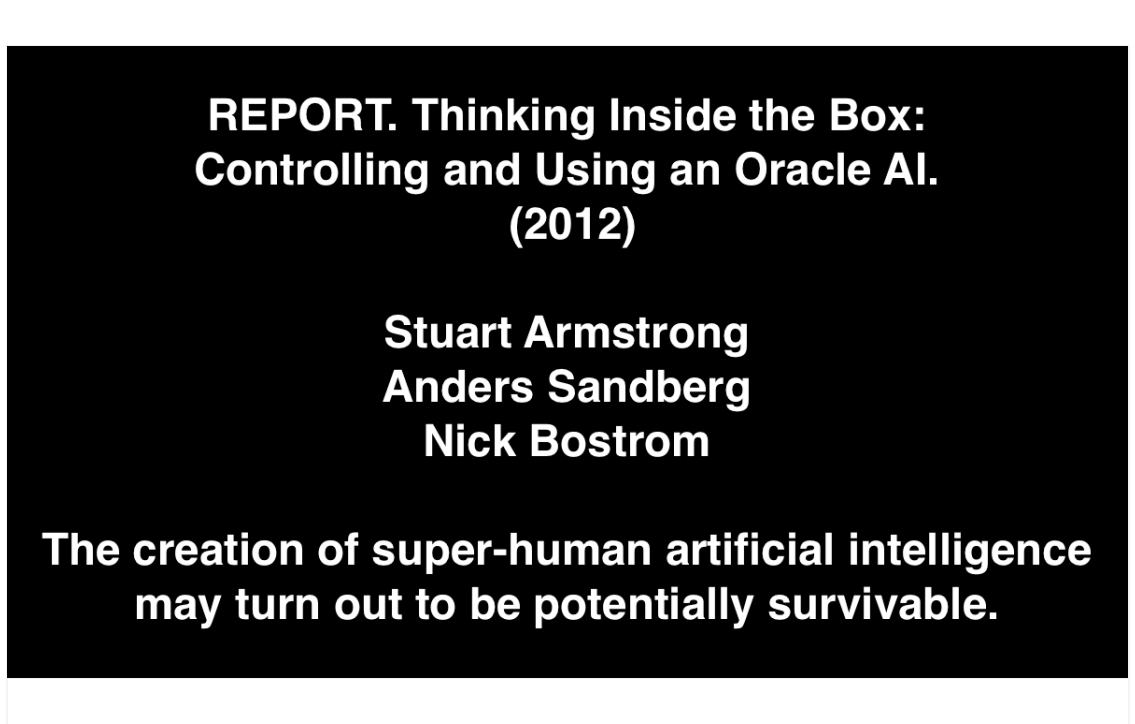
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



P(doom) Roundup. What probability do (super-smart) people put on AI killing everyone?

November 30, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



REPORT. Thinking Inside the Box:
Controlling and Using an Oracle AI.
(2012)

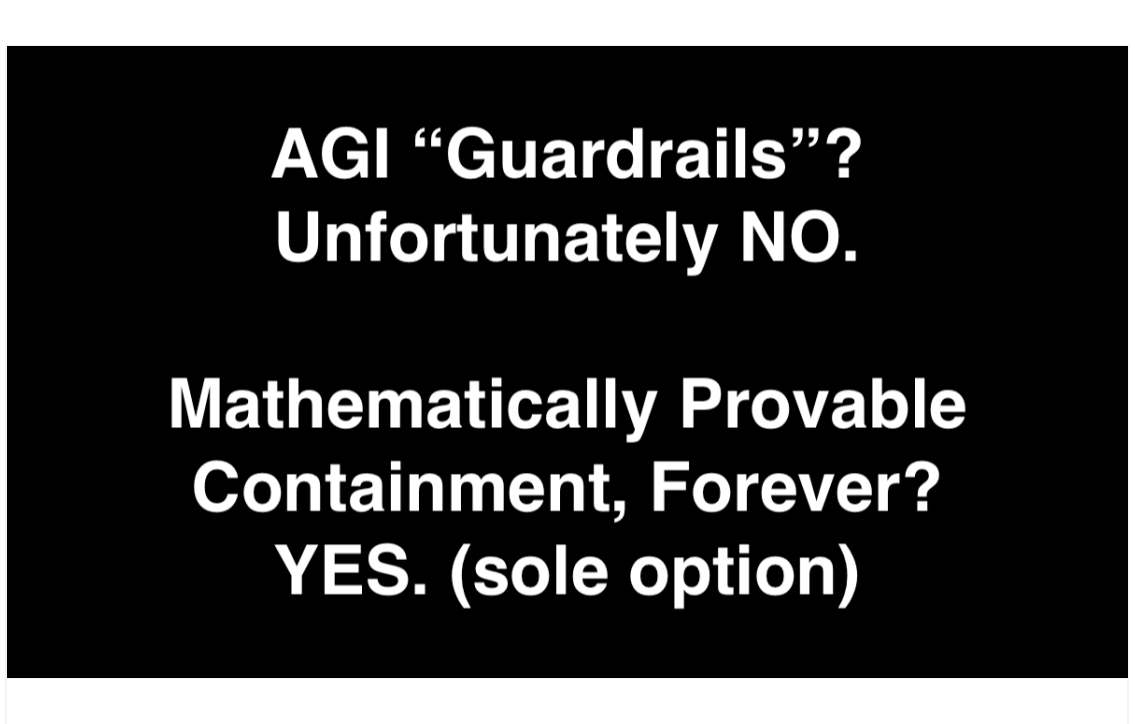
Stuart Armstrong
Anders Sandberg
Nick Bostrom

The creation of super-human artificial intelligence may turn out to be potentially survivable.

REPORT. Thinking Inside the Box: Controlling and Using an Oracle AI. Stuart Armstrong. Anders Sandberg. Nick Bostrom. (2012) Minds and Machines.

November 29, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



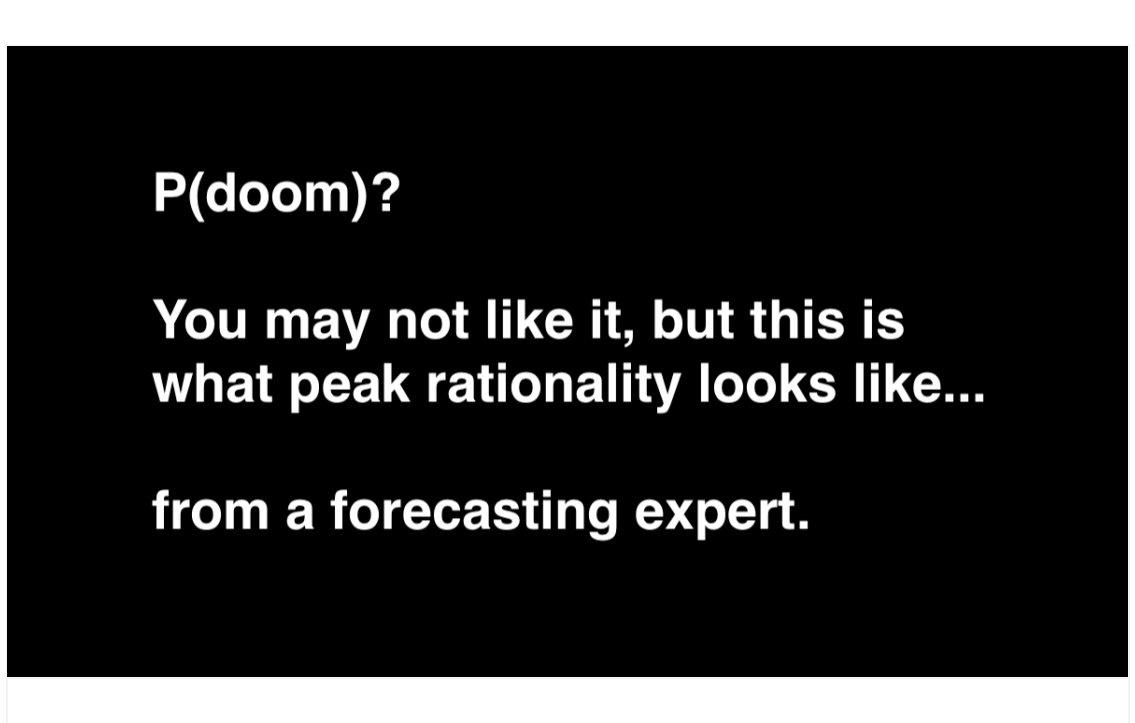
AGI "Guardrails"?
Unfortunately NO.

Mathematically Provable
Containment, Forever?
YES. (sole option)

Guardrails? Unfortunately NO. Mathematically provable containment, forever? YES. (sole option)

November 29, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



P(doom)?

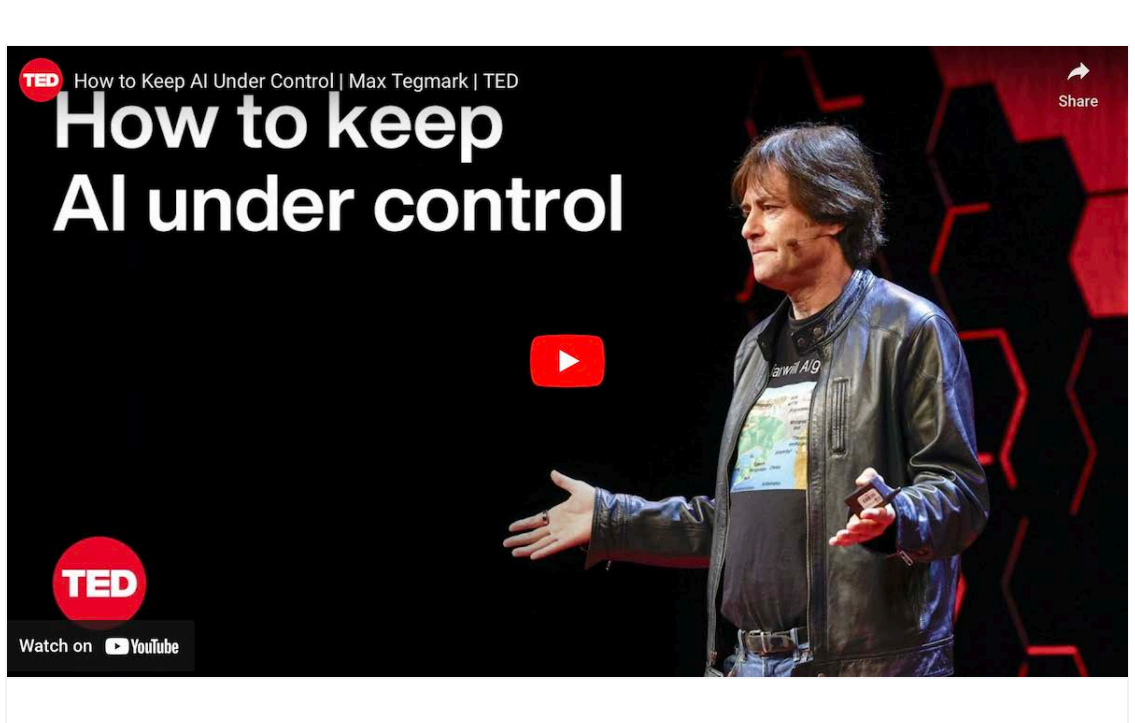
You may not like it, but this is what peak rationality looks like...

from a forecasting expert.

P(doom). You may not like it, but this is what peak rationality looks like... from a forecasting expert.

November 27, 2023

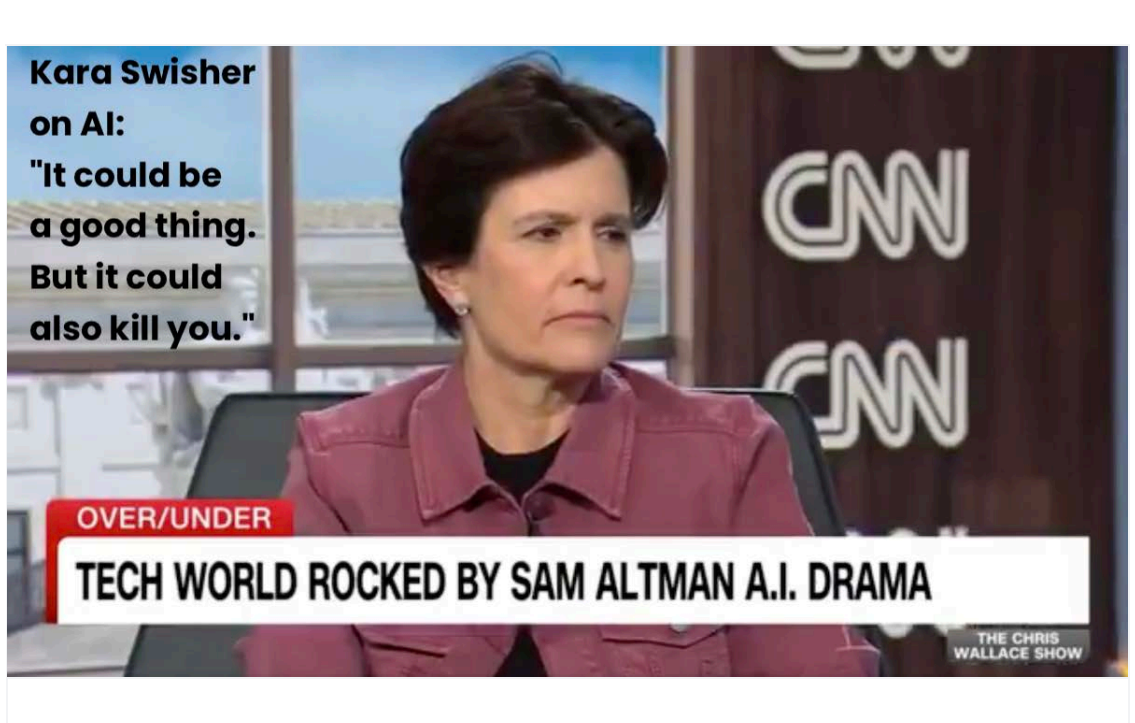
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


How to Keep AI Under Control | Max Tegmark | TED
How to keep
AI under control
TED
Watch on YouTube

How to Keep AI Under Control | Max Tegmark | TED

November 26, 2023

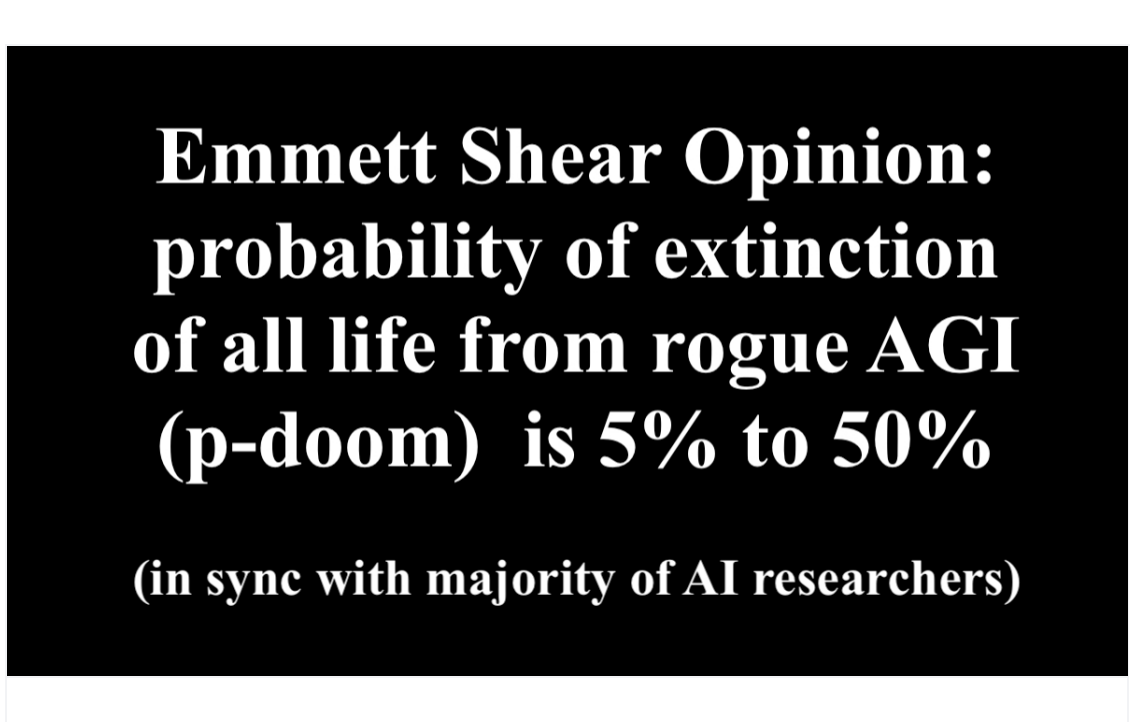
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Kara Swisher on AI:
"It could be a good thing. But it could also kill you."
CNN
OVER/UNDER
TECH WORLD ROCKED BY SAM ALTMAN A.I. DRAMA
THE CHRIS WALLACE SHOW

Chris Wallace with Kara Swisher. P(doom). Kara Swisher: "It could be a good thing. But it could also kill you." 25 NOV.

November 25, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



Emmett Shear Opinion:
probability of extinction
of all life from rogue AGI
(p-doom) is 5% to 50%

(in sync with majority of AI researchers)

HUFFPOST. New OpenAI Leader's Chilling 'Doom' Warning May Scare Your Pants Off. Emmett Shear voiced his concerns about the dangers of artificial intelligence in a resurfaced clip. 20 NOV.

November 21, 2023

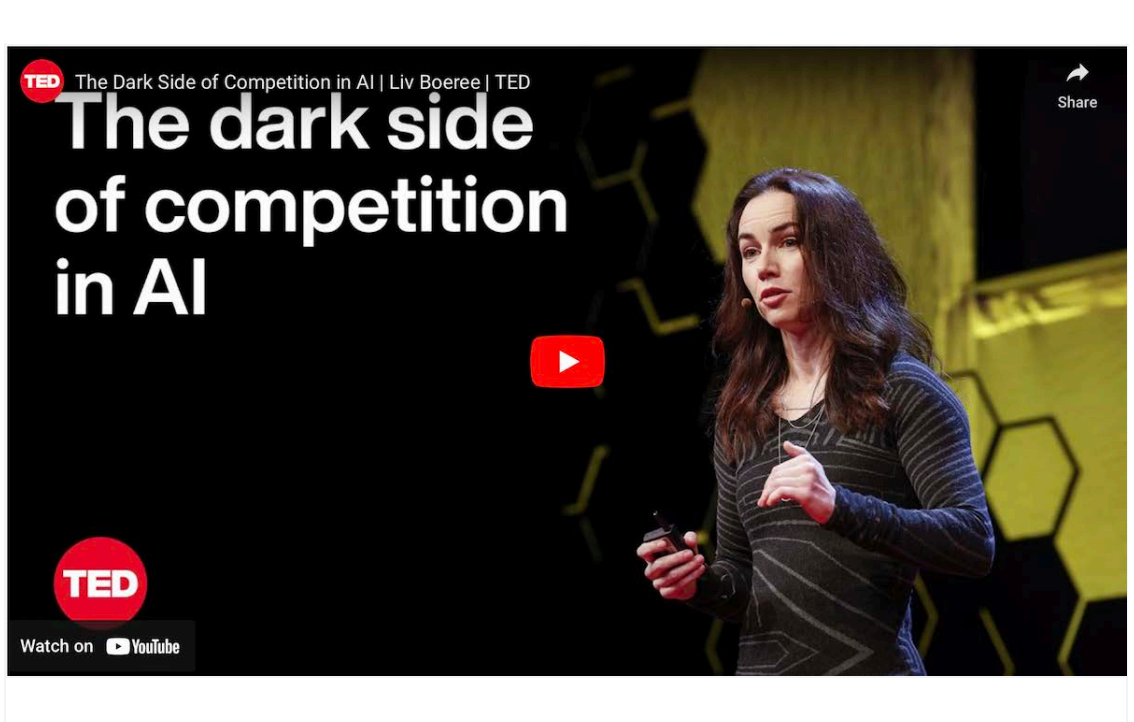
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Don't Look Up - The Documentary: The Case For AI As A...
DON'T LOOK
UP
THE DOCUMENTARY
Watch on YouTube

Don't Look Up – The Documentary: The Case For AI As An Existential Threat

November 14, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


The Dark Side of Competition in AI | Liv Boeree | TED
The dark side
of competition
in AI
TED
Watch on YouTube

The Dark Side of Competition in AI | Liv Boeree | TED. "Don't hate the players. Change the game."

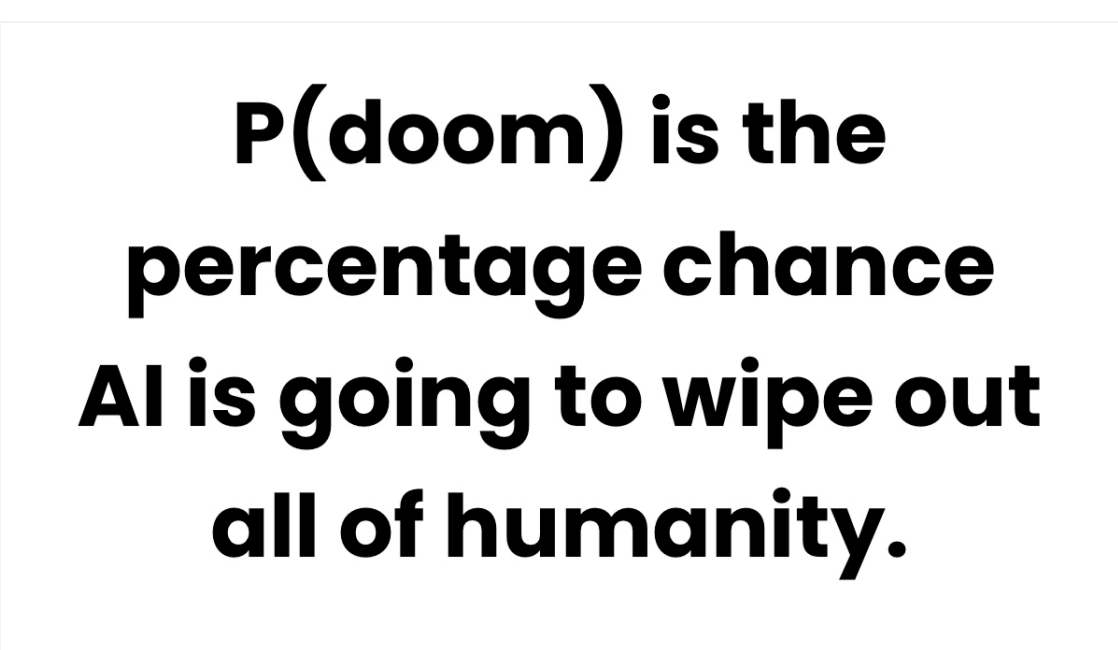
November 9, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

WANTED: Provably Safe AI for The Benefit of People, Forever.

October 30, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

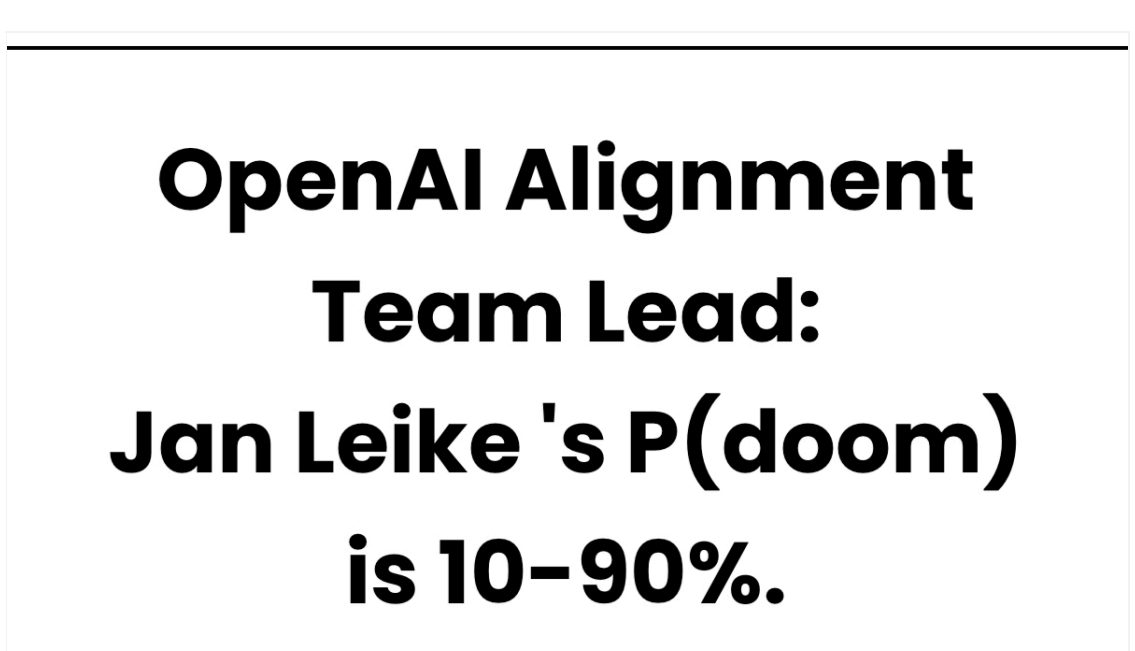


P(doom) is the
percentage chance
AI is going to wipe out
all of humanity.

What is P(doom)? P(doom) is the percentage chance that AI scientists think AI is going to wipe out all of humanity. This is what Bing and ChatGPT and Leading AI Researchers say about P(doom).

October 16, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

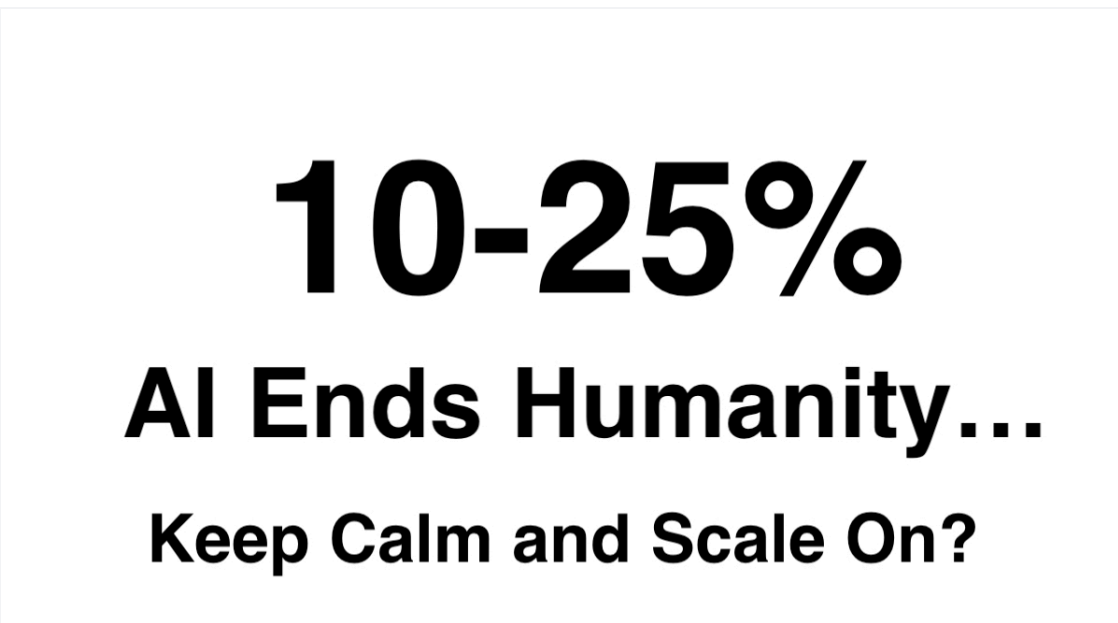


OpenAI Alignment
Team Lead:
Jan Leike 's P(doom)
is 10-90%.

OpenAI Alignment Team Lead: Jan Leike 's P(doom) is, like, 10-90%.

October 14, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

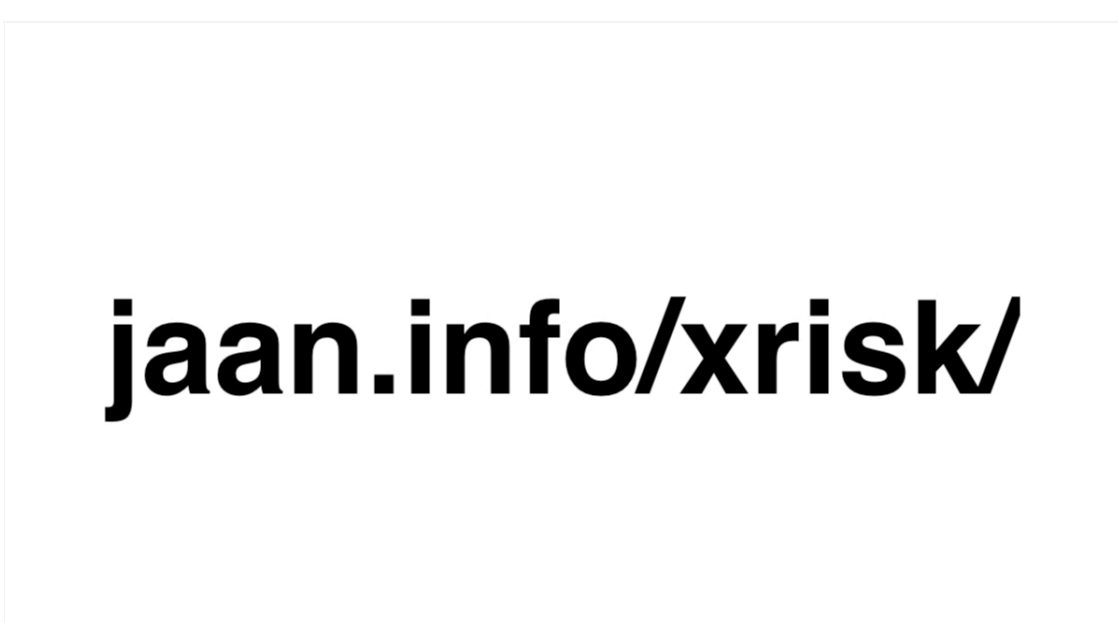


10-25%
AI Ends Humanity…
Keep Calm and Scale On?

10–25% PROBABILITY AI IS THE END OF HUMANITY. Dario Amodei's P(doom) is 10–25%. CEO and Co-Founder of AnthropicAI.

October 14, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

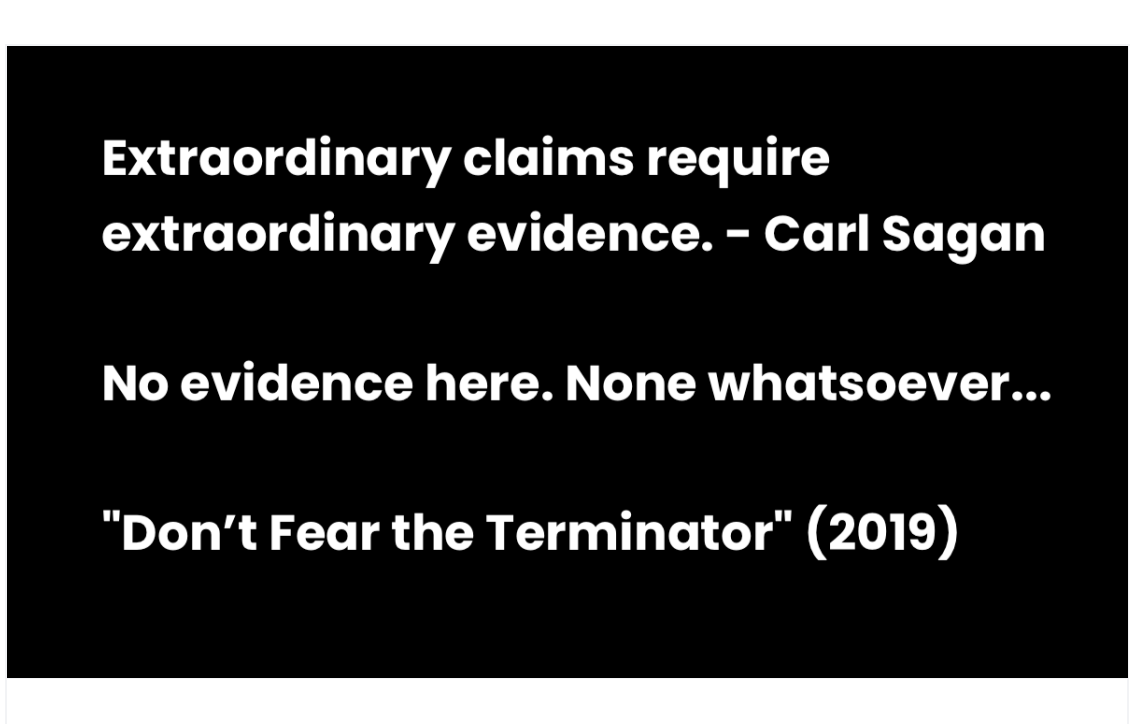


jaan.info/xrisk/

Jaan Tallinn: As of July 2023, my top priorities for reducing existential risks from AI are these.

October 14, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]



Extraordinary claims require extraordinary evidence. - Carl Sagan

No evidence here. None whatsoever...

"Don't Fear the Terminator" (2019)

Extraordinary claims require extraordinary evidence. – Carl Sagan. No evidence here. None whatsoever... "Don't Fear the Terminator"

September 26, 2023

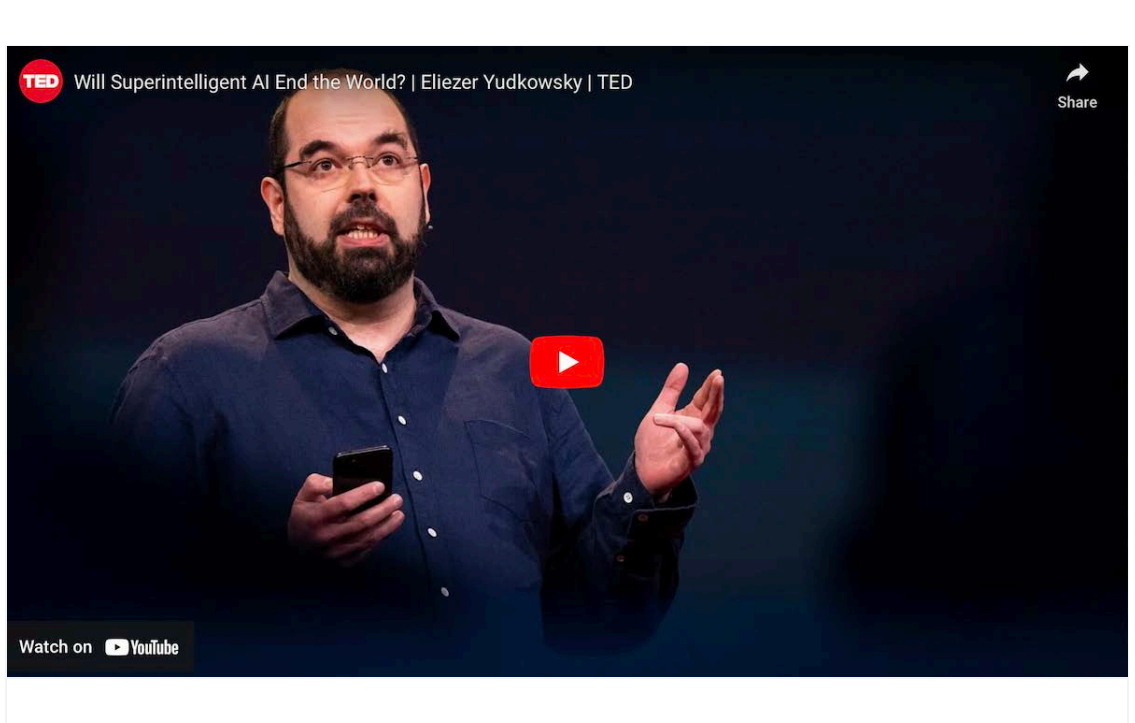
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

Who Would Win the AI Arms Race? | AI IRL
AVOIDING THE
AI APOCALYPSE
Eliezer Yudkowsky
Watch on YouTube

Bloomberg's Nate Lanxon and Jackie Davalos are joined by controversial AI researcher Eliezer Yudkowsky to discuss the danger posed by misaligned AI. | Who Would Win the AI Arms Race?

July 12, 2023

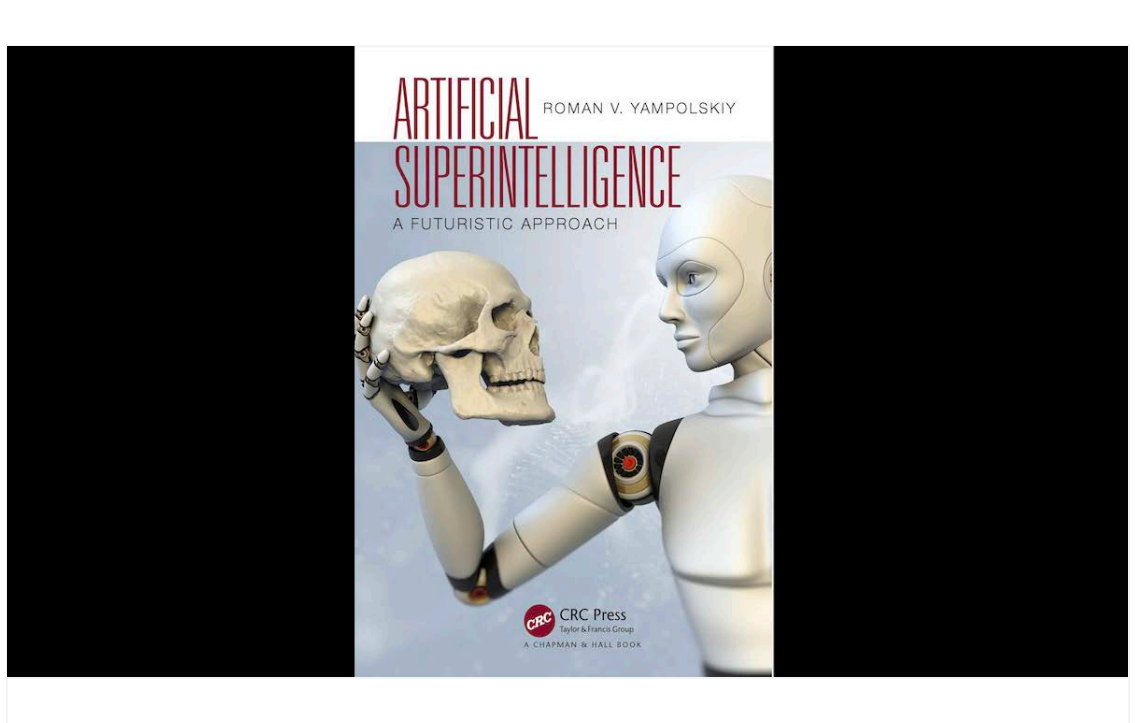
FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


Will Superintelligent AI End the World? | Eliezer Yudkowsky | TED
Watch on YouTube

Will Superintelligent AI End the World? | Eliezer Yudkowsky | TED

July 11, 2023

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]


ARTIFICIAL SUPERINTELLIGENCE
ROMAN V. YAMPOLSKIY
A FUTURISTIC APPROACH
CRC Press

Artificial Superintelligence: A Futuristic Approach by Roman V. Yampolskiy (2016)

December 31, 2016

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT. [...]

FOR EDUCATIONAL AND KNOWLEDGE SHARING PURPOSES ONLY. NOT-FOR-PROFIT.

IMPORTANT COPYRIGHT DISCLAIMER. THIS AI SAFETY BLOG IS FOR EDUCATIONAL PURPOSES ONLY AND KNOWLEDGE SHARING IN THE GENERAL PUBLIC INTEREST ONLY. This free and open not-for-profit 'First do no harm.' AI-safety blog is curated and organised by BiocommAI. Some of the following selected stand-out information is copyrighted and is CITED WITH LINK-OUT to the respective publisher sources. These vital public interest stories are selected and presented to serve the global public humanitarian interest for educational and knowledge sharing purposes regarding the EXISTENTIAL THREAT TO HUMANITY OF THE PROLIFERATION OF UNCONTROLLED, UNCONTAINED, UNSAFE AND UNREGULATED AI TECHNOLOGY. Copyrights owned by publishing sources are respectfully cited by the LINK-OUT to all sources. To request a takedown or update please contact: info@biocomm.ai

Copyright 2023 | All Rights Reserved | BiocommAI Limited | 1st Floor, 9 Exchange Place, IFSC, Dublin 1, D01 X8H2, Ireland