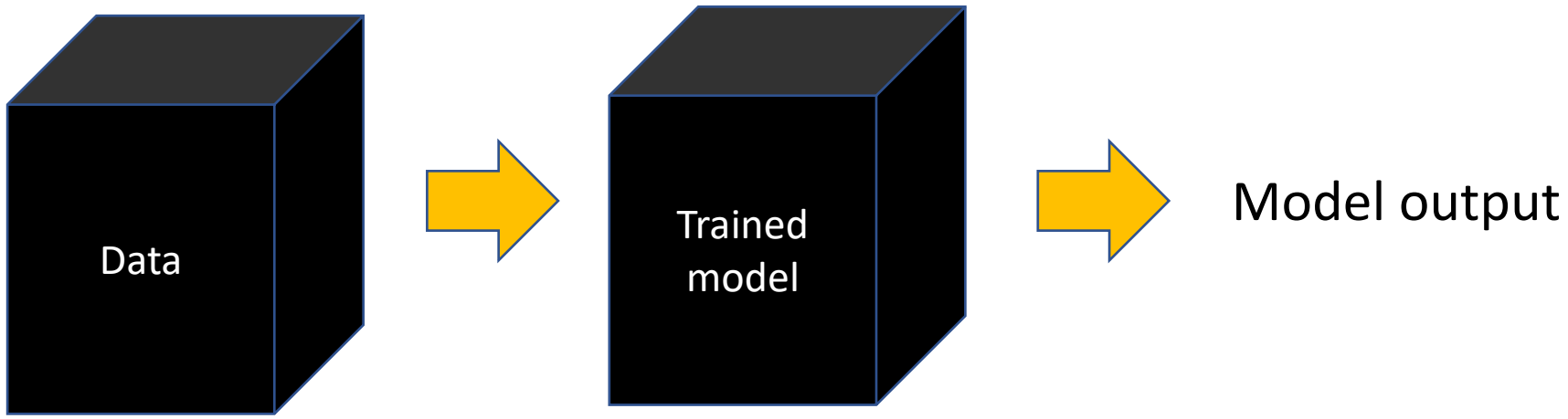
Societal, ethical and legal implications of AI

Sonja Aits
Lund University, 20250924

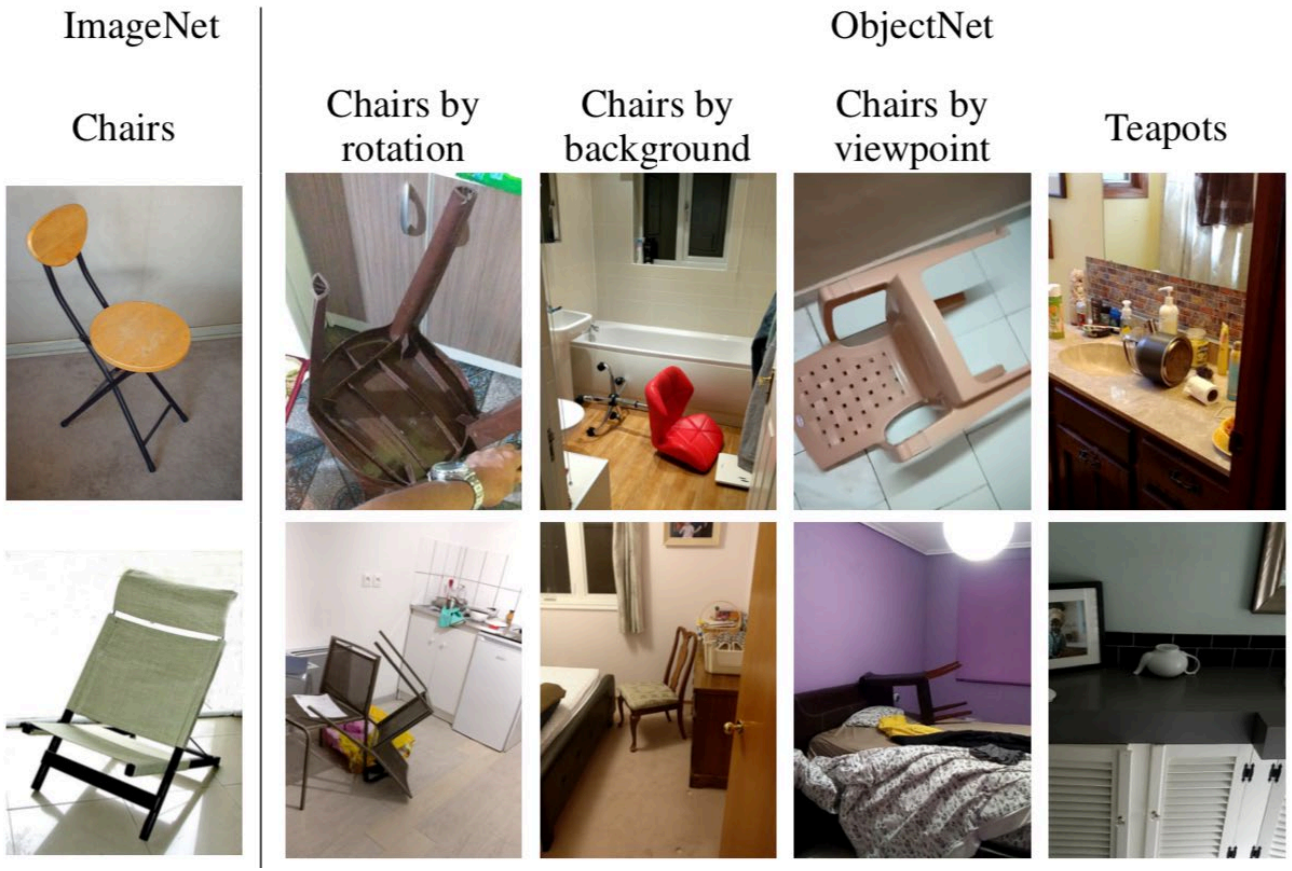
Is AI a means or obstacle in the sustainable transformation?

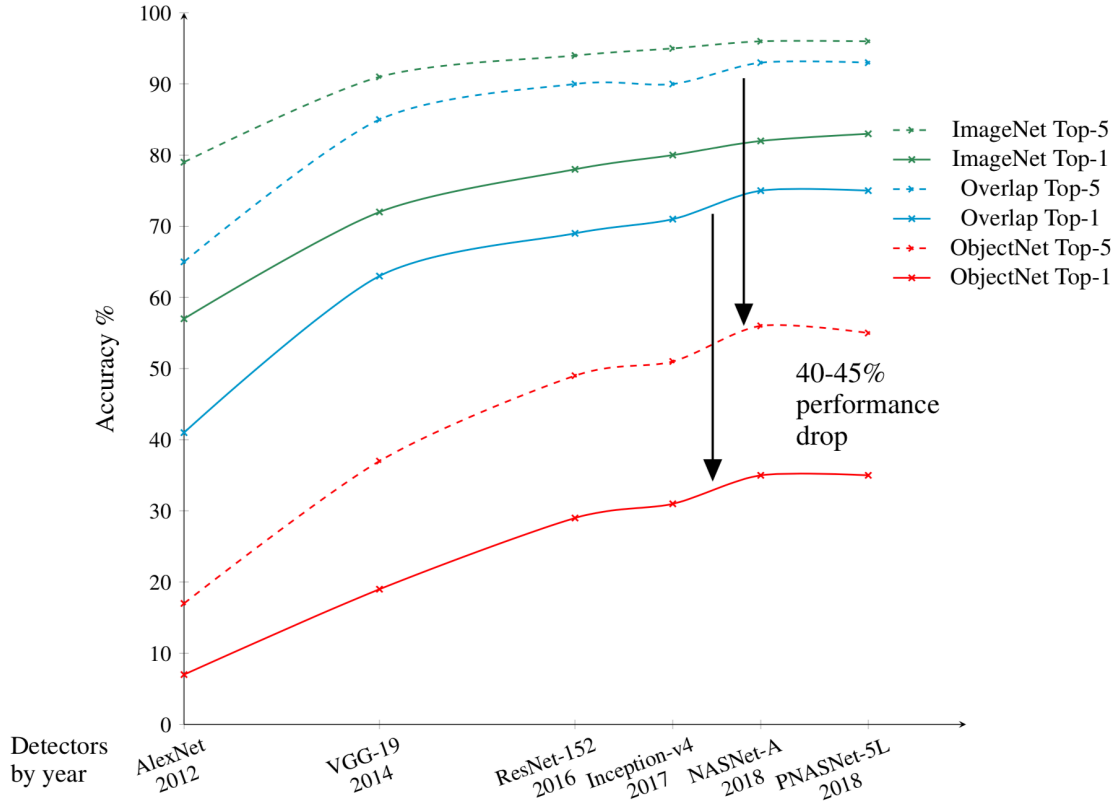
Can we trust AI models?

Deep learning models lack transparency
“Black box issue”



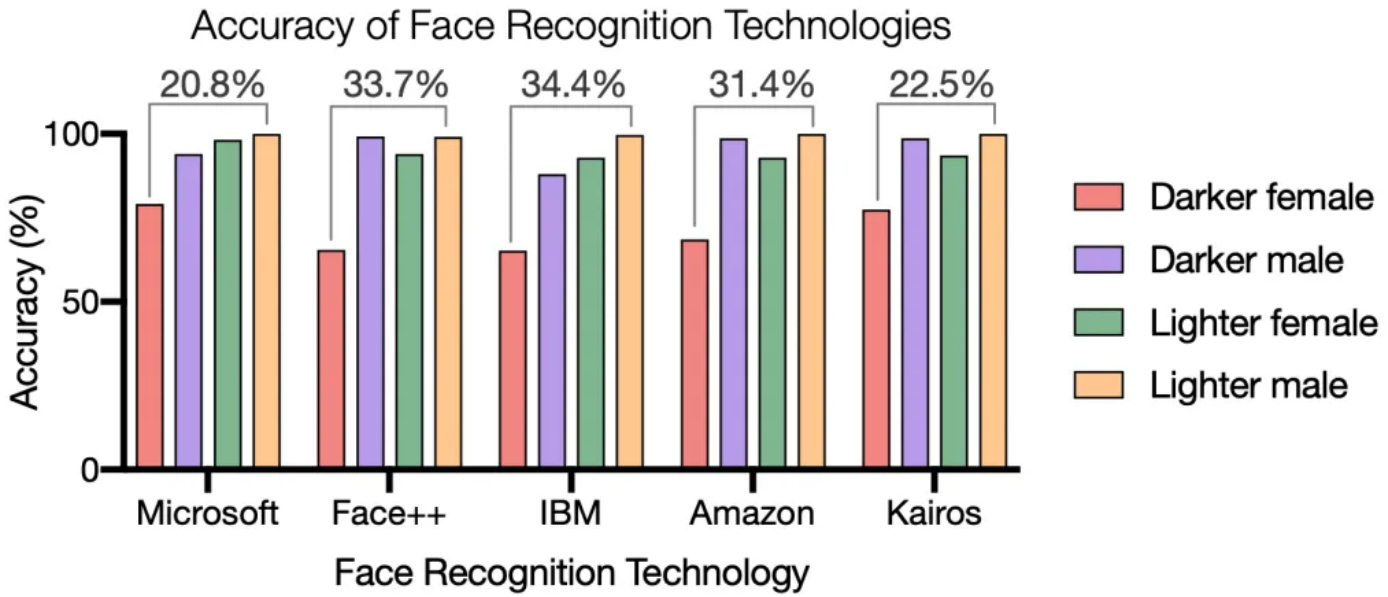
“Hidden” biases in the training data affect model performance





Barbu et al. Advances in Neural Information Processing Systems 32, pages 9448–9458. 2019.

Biased training data makes AI models prejudiced



https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

SHARE

REPORTS | PSYCHOLOGY

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan[1,*], Joanna J. Bryson[1,2,*], Arvind Narayanan[1,*]
+ See all authors and affiliations

Science 14 Apr 2017:
Vol. 356, Issue 6334, pp. 183-186
DOI: 10.1126/science.aal4230

MICROSOFT | WEB | TL;DR

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent | Mar 24, 2016, 6:43am EDT
Via The Guardian | Source TayandYou (Twitter)

https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Incorrect medical AI can cause real harm

AI can be prejudiced or simply wrong

Report: IBM Watson delivered 'unsafe and inaccurate' cancer recommendations

JULY 25, 2018 BY FINK DENSFORD — LEAVE A COMMENT

"Garbage in – Garbage out"

Privacy concerns

REGULATION

There's No Such Thing as Anonymous Data

by Scott Berinato

FEBRUARY 09, 2015

Anonymous patient data may not be as private as previously thought

REUTERS By Linda Carroll, Reuters • December 21, 2018

Experts warn of possible unintended use of DNA amid 23andMe bankruptcy

Washington Post technology columnist Geoffrey Fowler tells CNN's Laura Coates how DNA submitted to 23andMe may wind up being used for other purposes if the data is sold during the company's bankruptcy.

02:36 - Source: CNN

svt NYHETER Nyheter Lokalt Sport SVT Play Barn

/ INRIKES

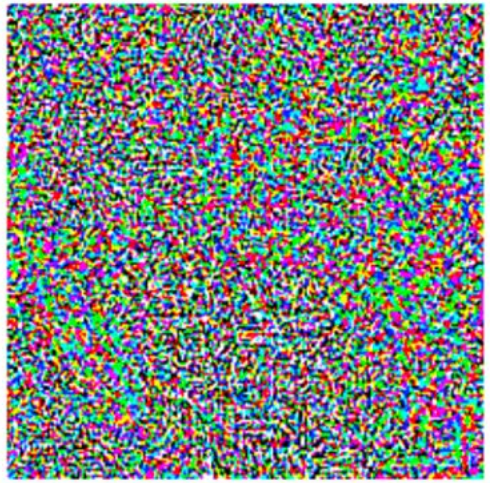
2,7 miljoner inspelade samtal till 1177 Vårdguiden helt oskyddade på internet

Who is responsible when things go wrong with AI?

AI infrastructure is vulnerable to malicious attacks

AI tools are vulnerable to malicious attacks

$+ .007 \times$



$=$

"panda"

57.7% confidence

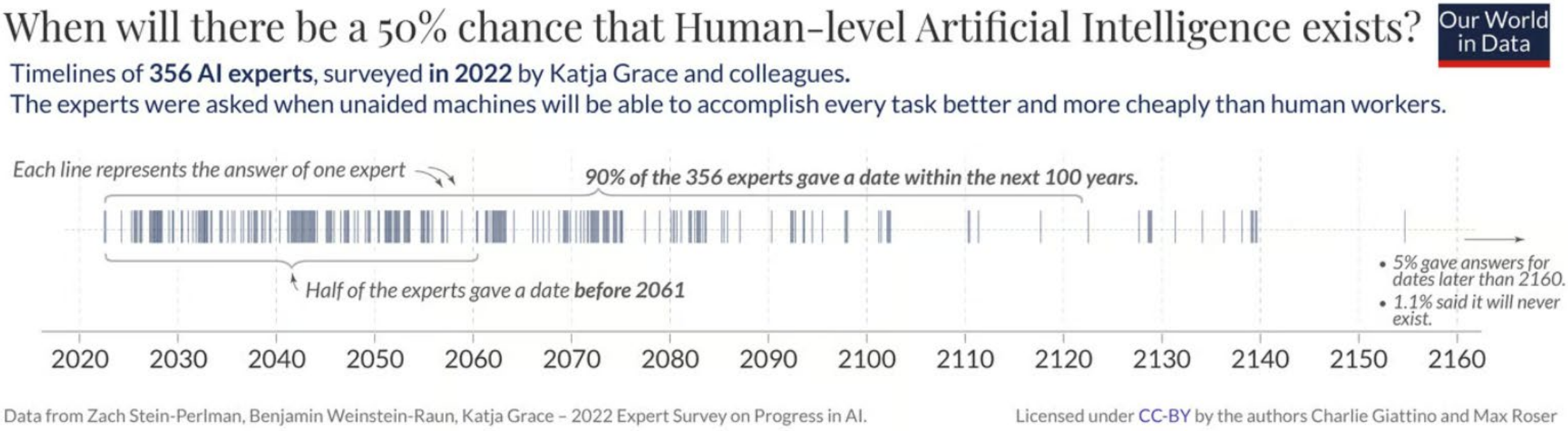
noise

"gibbon"

99.3% confidence

Goodfellow et al, ICLR 2015

Will we have Artificial General Intelligence soon?



Do you support open research?

What could you do with AI?

What could you do with AI if you were?

- the leader of an EU country
- the head of the WHO
- a primary care doctor in a developing nation
- a specialist in the world’s best hospital
- the head of Amnesty International

TECH

Microsoft health unit is a ‘multi-billion dollar business,’ medical chief says

PUBLISHED THU, JUL 26 2018 • 1:36 AM EDT

Facebook Ventures Into Medical Imaging Research

BY SHOBHIT SE

TECH • APPLE

Apple’s Healthcare Take Could Be $313 Billion by 2027, Analysts Say

SCIENCE BUSINESS TECH

Google's project to 'cure death,' Calico, announces $1.5 billion research center

67

Calico will work with big pharma to create and market new drugs for diseases like cancer and Alzheimer's

By Ben Popper | Sep 3, 2014, 2:33pm EDT

Via Re/code | Source Calico Labs

BUS

Column: Alexa may be key to Amazon’s looming domination of the healthcare market

What could you do with AI if you were?

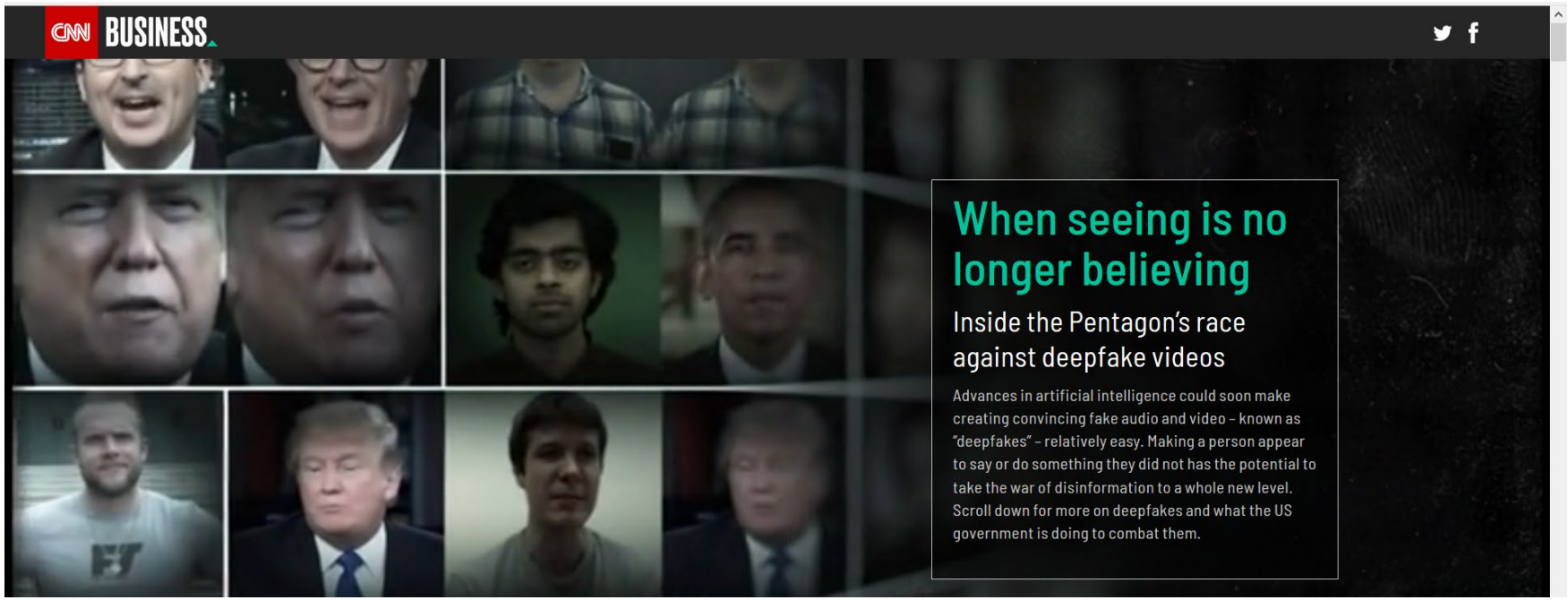
- the leader of an EU country
- the head of the WHO
- a primary care doctor in a developing nation
- a specialist in the world's best hospital
- the head of Amnesty International
- the leader of a global tech company

TECH \ ARTIFICIAL INTELLIGENCE

OpenAI has published the text-generating AI it said was too dangerous to share

The lab says it's seen 'no strong evidence of misuse so far'

By James Vincent | Nov 7, 2019, 7:24am EST



“Deep Fakes” produced by generative AI can be abused in many ways

Many different data types can be generated: video, images, sounds, etc.

Legitimate applications	Illegitimate applications
Model training	Fraud
Research with simulations	Blackmail
Education	Defamation
Entertainment	Misinformation

China Uses DNA to Track Its People, With the Help of American Expertise

The Chinese authorities turned to a Massachusetts company and a prominent Yale researcher as they built an enormous system of surveillance and control.

Access more for free.

COLLAPSE

The New York Times

Create a free account or log in to access more of The Times.

CONTINUE

What could you do with AI if you were?

- the leader of an EU country
- the head of the WHO
- a primary care doctor in a developing nation
- a specialist in the world's best hospital
- the head of Amnesty International
- the leader of a global tech company
- a terrorist
- a dictator
- a white supremacist
- a criminal

Do you think AI application in your research domain fall under dual use?

Should we have open research in the field of AI?

AI **for** sustainability

Sustainability **of** AI

Water

Energy

Raw materials

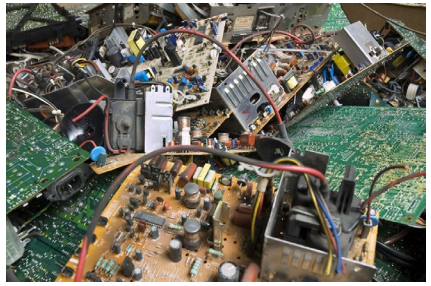
Production

Transport

Supercomputer

Researcher

Waste and pollution



State-of-the-art models cost > 100 million dollars to train
→ Field dominated by a few large tech companies

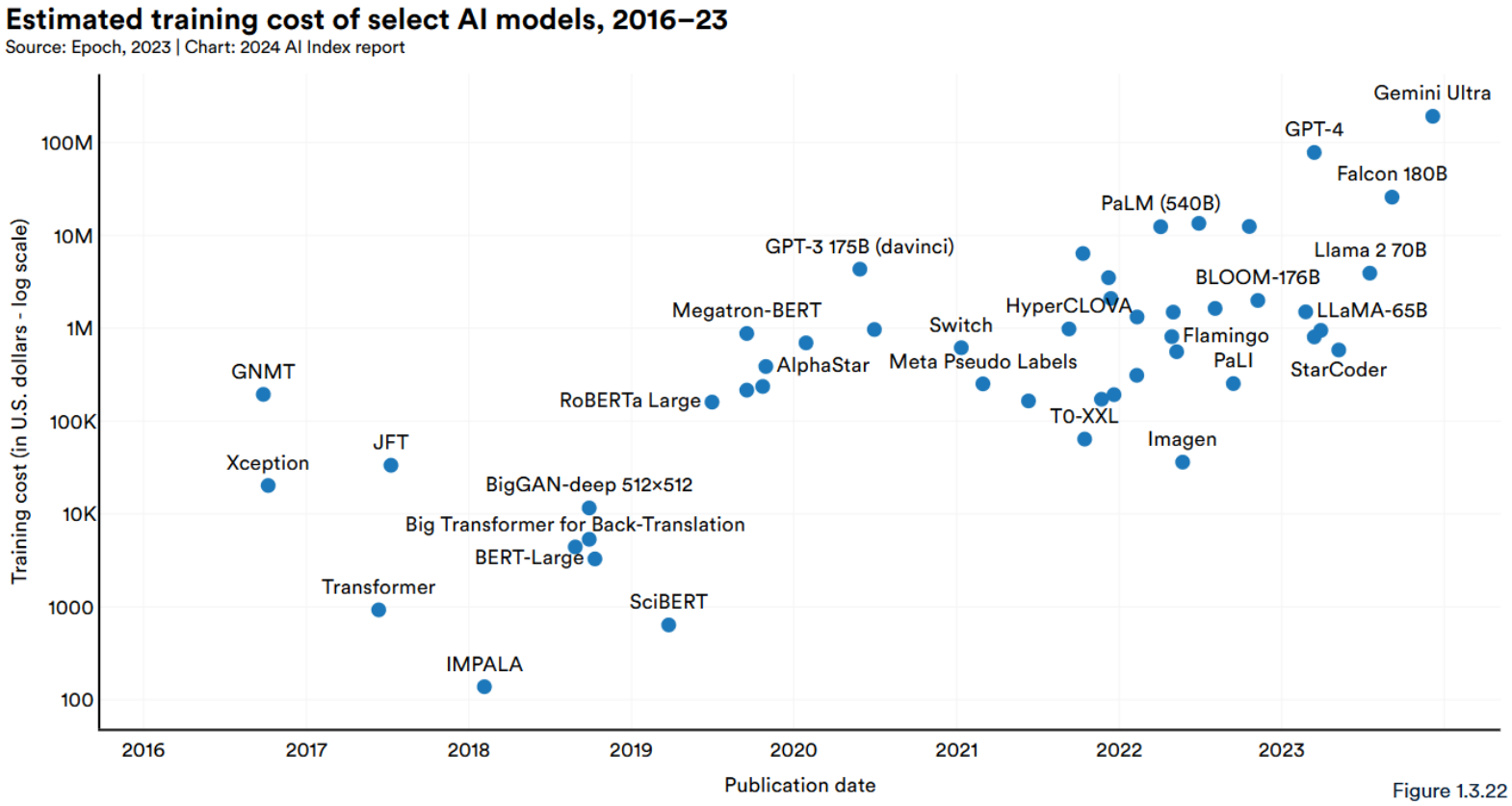


Figure 1.3.22

Source: https://arxiv.org/pdf/2405.19522

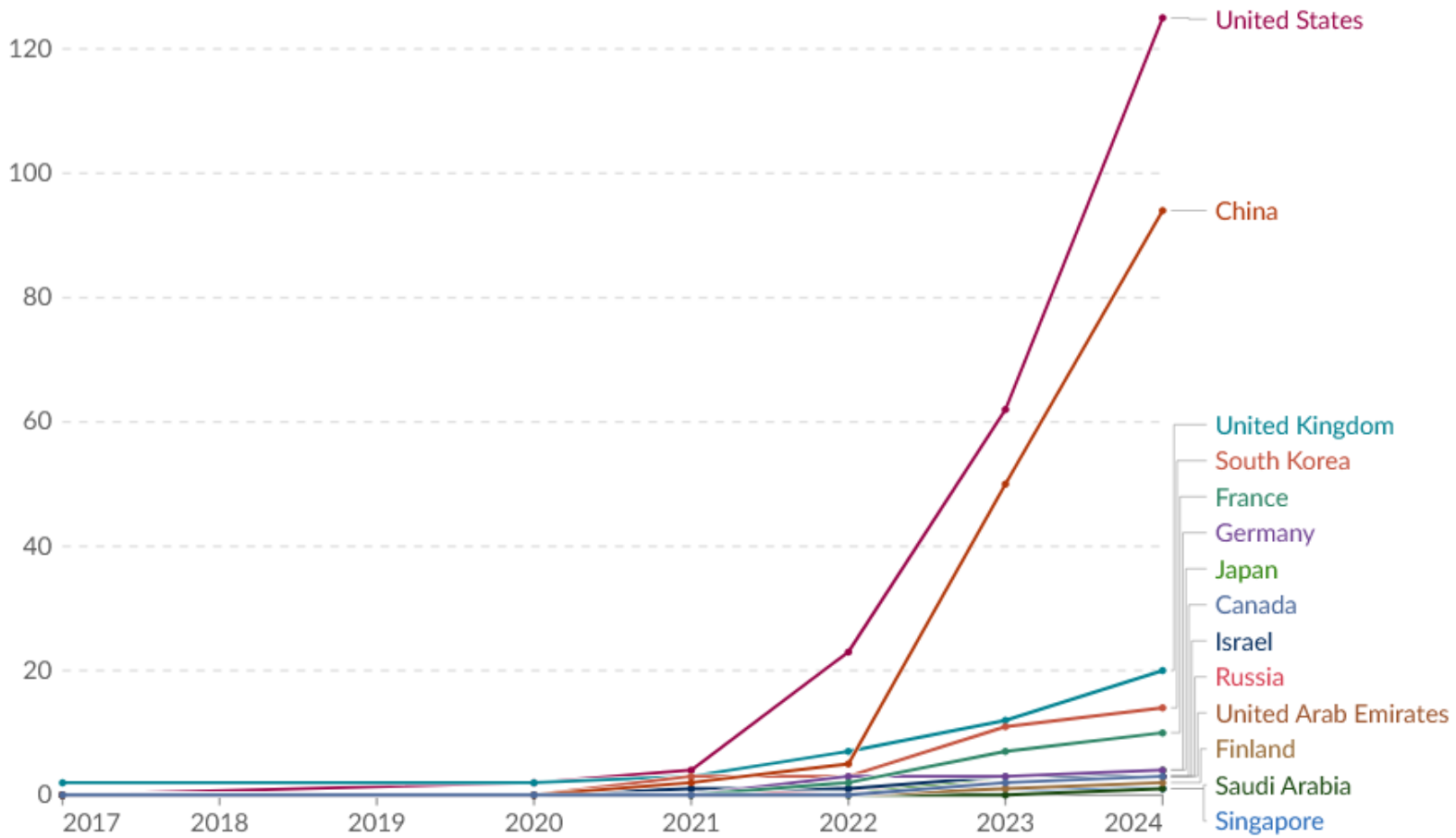
Cumulative number of large-scale AI systems by country since 2017

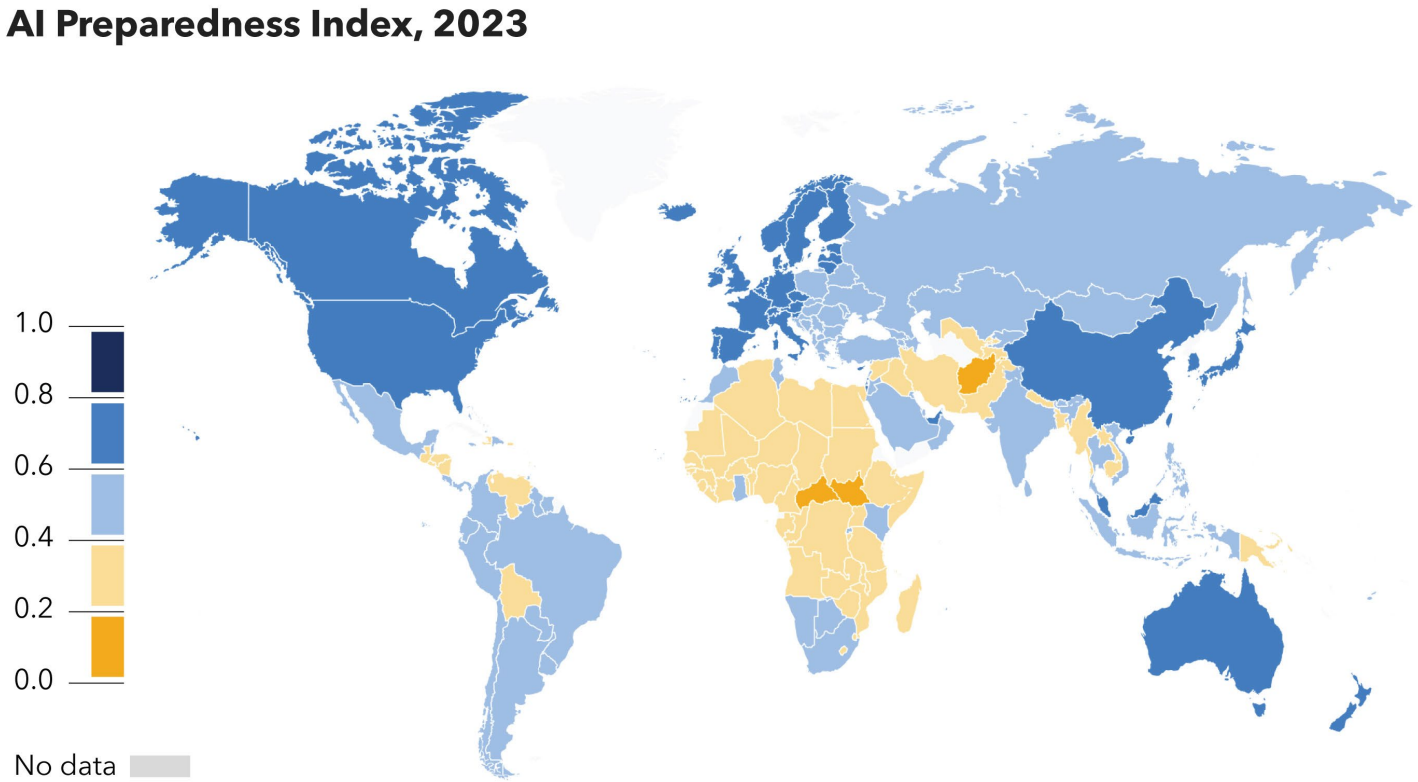
Refers to the location of the primary organization with which the authors of a large-scale AI systems are affiliated.

Our World in Data

Table | Chart

Edit countries and regions

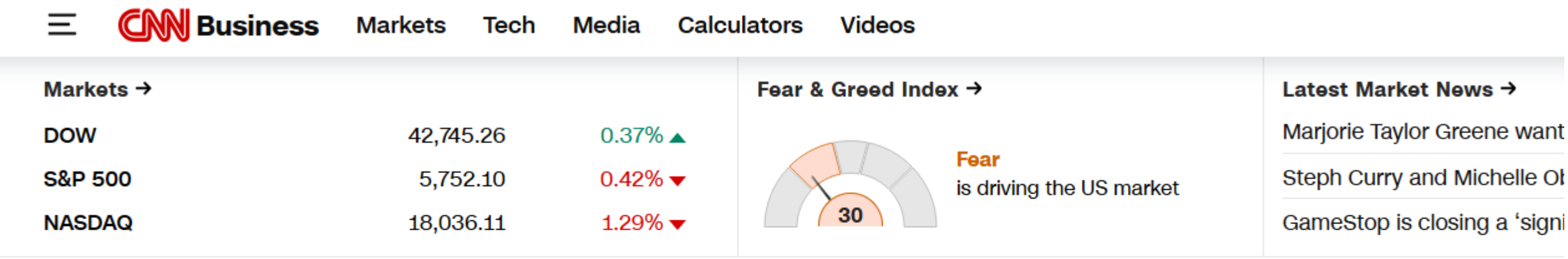




Source: IMF
https://www.imf.org/-/media/Images/IMF/Data/Home/Rotator/ai-index-map-lp-1.ashx

Should we have open research in the field of AI?

AI will take over many types of tasks and jobs

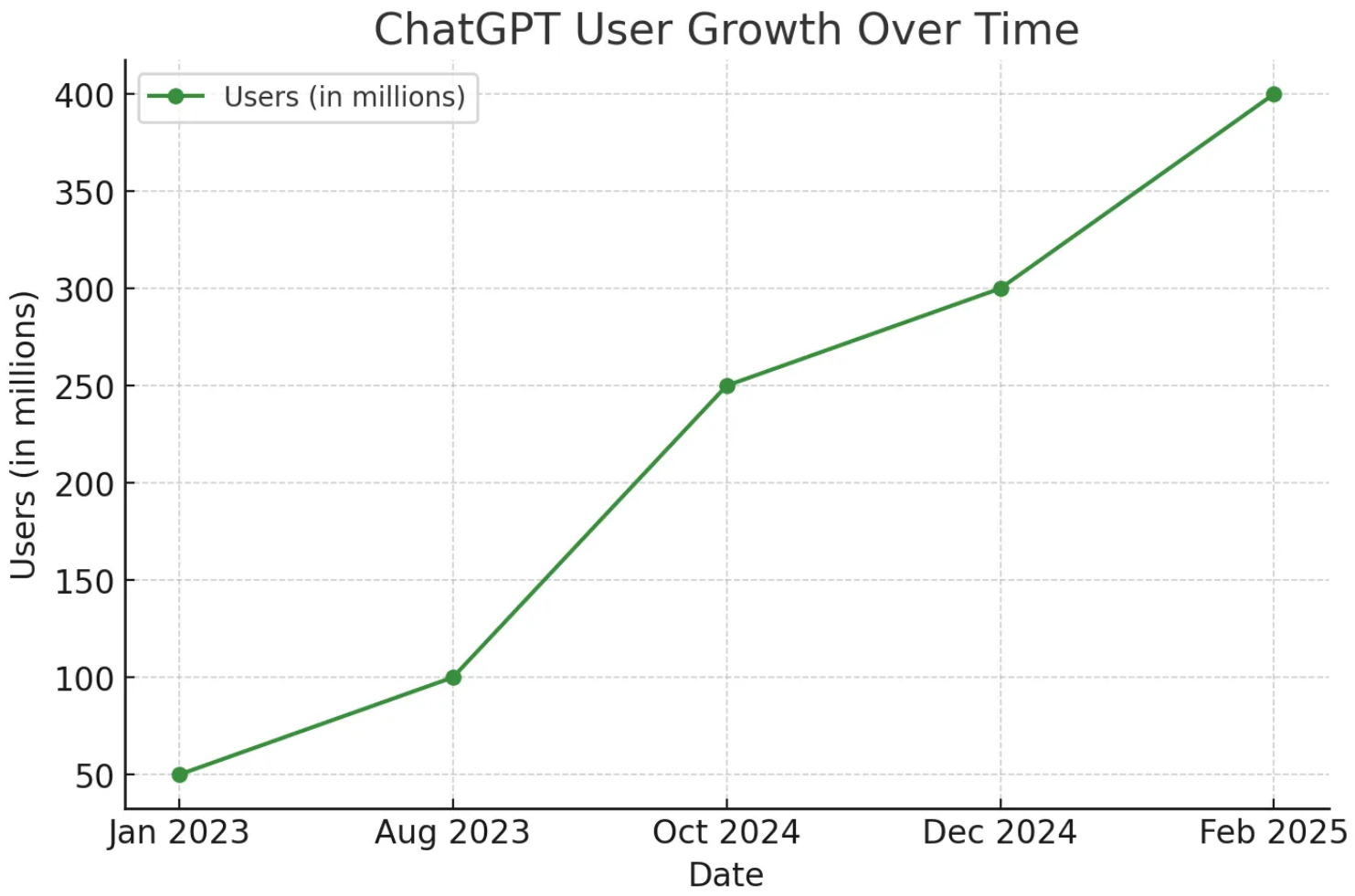


41% of companies worldwide plan to reduce workforces by 2030 due to AI

By Olesya Dmitracova, CNN

2 minute read · Published 7:57 AM EST, Wed January 8, 2025

Generative chatbots shape the way we access information



1 billion queries every day

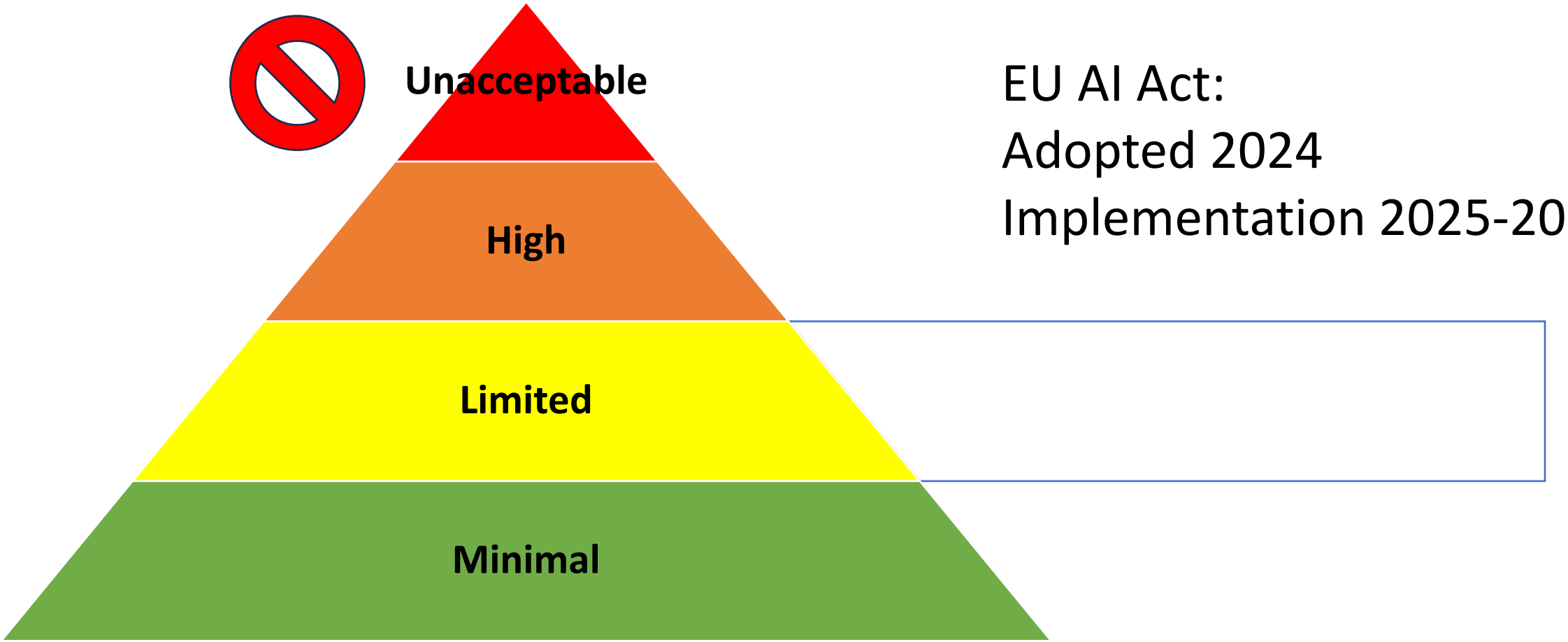
https://nerdynav.com/chatgpt-statistics/

7 requirements for ethical AI

- Human agency and oversight
- Technical Robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

AI regulation is struggling to keep up with rapid technology development



EU AI Act:
Adopted 2024
Implementation 2025-2027

Explainable AI

= Attempts to reduce the “black box issue” - make AI decisions **transparent, understandable, and trustworthy** for humans

- Transparency about various development steps (data, code, etc.)
- Investigating importance of input features
- Simpler models

Will AI make the world more sustainable?

Group discussion

Ethical, legal, environmental and societal issues related to AI

Discussion case:
Large language models and AI chatbots

AI chatbots (e.g. ChatGPT, Gemini) answer questions on a very large variety of topics and generate text, code and images. There are many beneficial applications for both society and science. The underlying models are trained on extremely large collections of websites, scientific literature, books and other inputs. They are then refined by humans who give feedback on the answers, both employees (typically in low-income countries) and end-users. Training data, procedures, and model details are mostly not made publicly available. Even if details were available the development of these tools cannot easily be replicated as it requires extensive resources. Access to the best models is typically provided for a monthly fee whereas slightly less capable models are available for free but with input data from users typically being used for further development.

Discuss the ethical, legal, environmental and societal issues for each case

- Ownership of the training data, AI tools and AI output
- Sources and risks of bias
- Transparency regarding development and output
- Societal benefits and negative impact (e.g. on the job market)
- Environmental benefits and negative impact
- Accessibility (e.g. for non-English speakers, low-income countries, people with handicaps)
- Possibilities for abuse by malicious actors
- Responsibility for correctness and safety
- Possibilities for oversight by public authorities and by the end-user
- Consequences of forbidding the development and usage of this type of AI in the EU

Debate format

- Pro statements
- Con statements