Cluster of Excellence
The Politics of Inequality

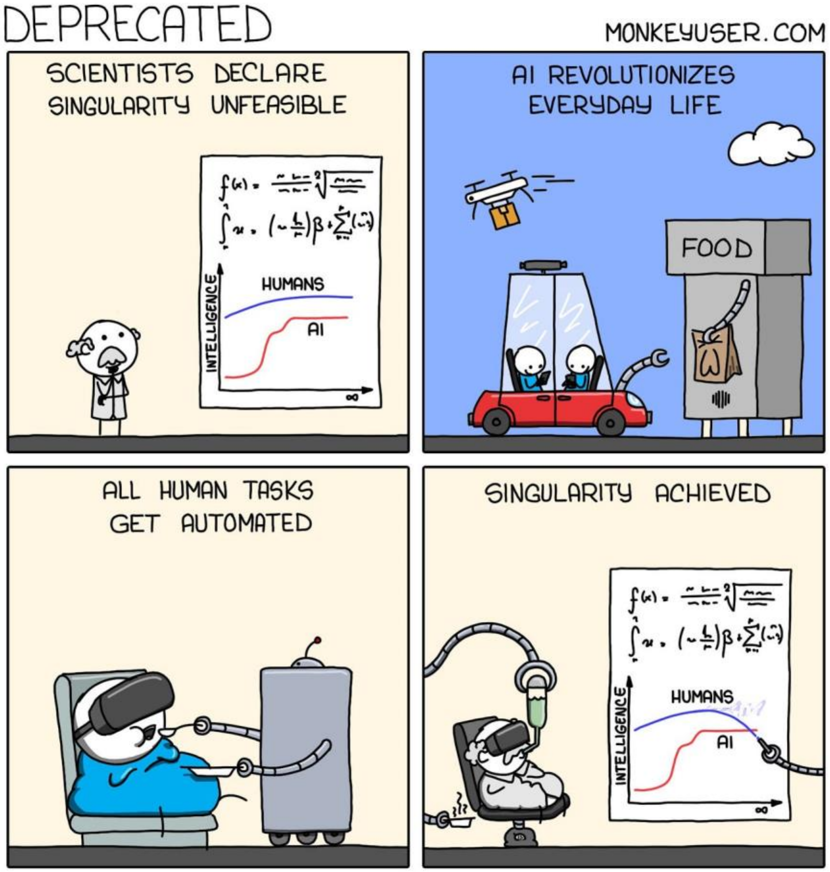
Steinbeis Transfer Center
Linguistic Data Analysis

UNIVERSITÄT PASSAU

Week 11:
AI, Ethics & Superintelligence

Annette Hautli-Janisz, Prof. Dr.

23 January 2025


DEPRECATED
MONKEYUSER.COM
SCIENTISTS DECLARE SINGULARITY UNFEASIBLE
HUMANS
AI
INTELLIGENCE
AI REVOLUTIONIZES EVERYDAY LIFE
FOOD
ALL HUMAN TASKS GET AUTOMATED
SINGULARITY ACHIEVED
HUMANS
AI
INTELLIGENCE

A couple of notes on the exam

- Learn the concepts/terminology/definitions and how to apply them (What is a morpheme/constituent?, What is inter-annotator agreement?, etc.)
- Learn to compute scores/similarities/probabilities etc. for retrieval examples

Bring a calculator!!

Today

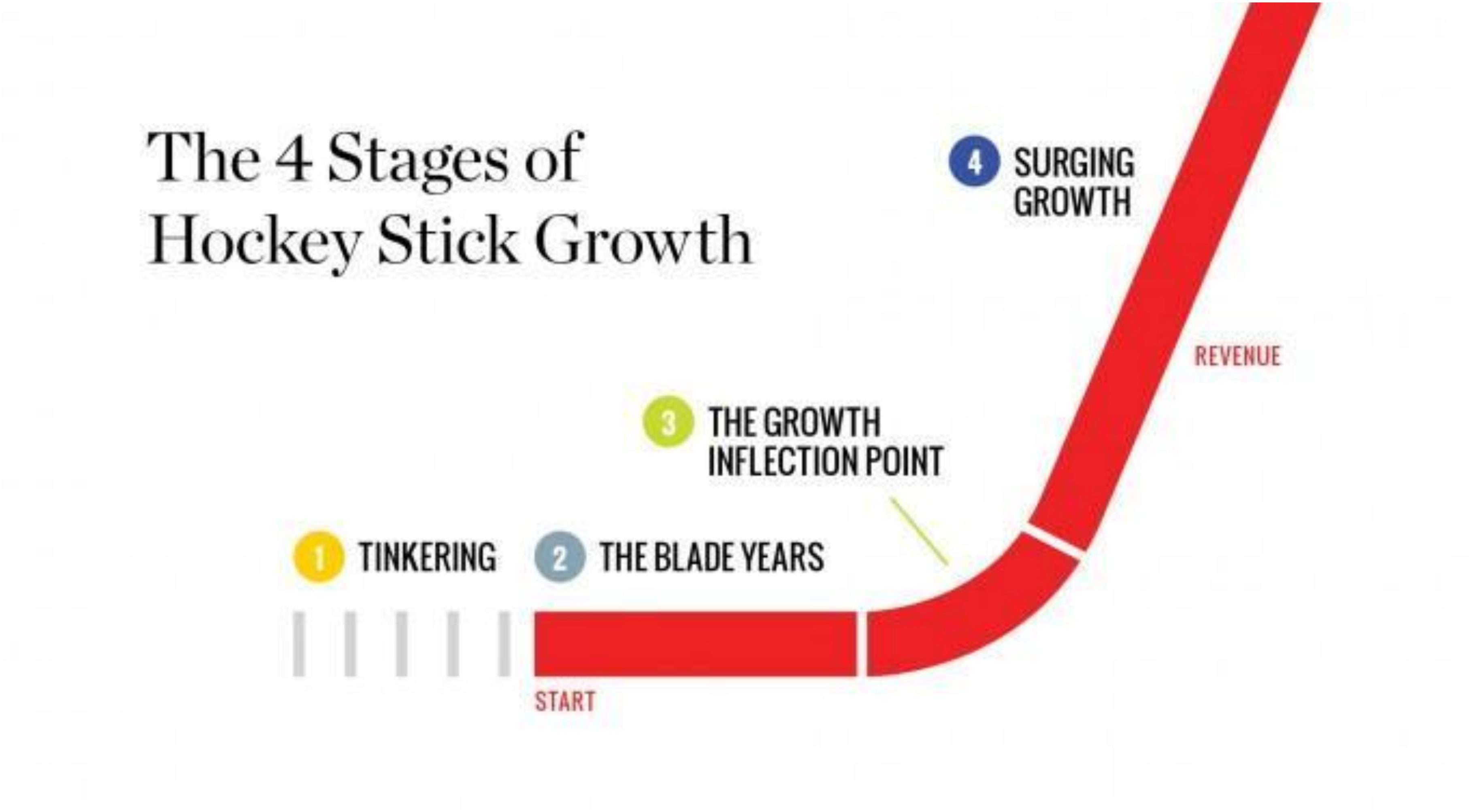
Part 1: Artificial Intelligence

- A brief history
- Strong versus weak AI

Part 2: AI, ethics and regulation

Part 2: Superintelligence

AI and the rate of growth


The 4 Stages of Hockey Stick Growth
4 SURGING GROWTH
REVENUE
3 THE GROWTH INFLECTION POINT
1 TINKERING
2 THE BLADE YEARS
START

AI and the rate of growth

A mere few million years ago:

- “We” were still swinging from the branches in the African canopy.
- The rise of Homo sapiens from our last common ancestor with the great apes happened swiftly.
- Upright posture, opposable thumbs and some relatively minor changes in brain size
  - Leap in cognitive ability.
  - Humans can think abstractly, communicate complex thoughts, culturally accumulate information over the generations far better than any other species on the planet.

AI and the rate of growth

12.000 years ago: Adoption of agriculture.

- Population densities rose along with the total size of the human population.
- More people → more ideas.
- Higher densities → ideas spread more easily.

- Some individuals develop specialized skills.

- Increases economic productivity and technological capacity.

Early “AI”

Talos of Crete

https://www.ancient-origins.net/myths-legends/talos-crete-00157

Early “AI”

Robots guarded Buddha’s relics in a legend of ancient India

March 13, 2019 10.40am GMT

Two small figures guard the table holding the Buddha’s relics. Are they spearmen, or robots? British Museum, CC BY-NC-SA

http://theconversation.com/robots-guarded-buddhas-relics-in-a-legend-of-ancient-india-110078

King Ajatasatru (reigned 492 to 460 B.C.) was famous for commissioning new military inventions.

Blue prints for Robots supposedly stolen from Rome.

Early “AI”

https://ancientcelebration.blogspot.com/2011/03/grand-procession-of-ptolemy_24.html

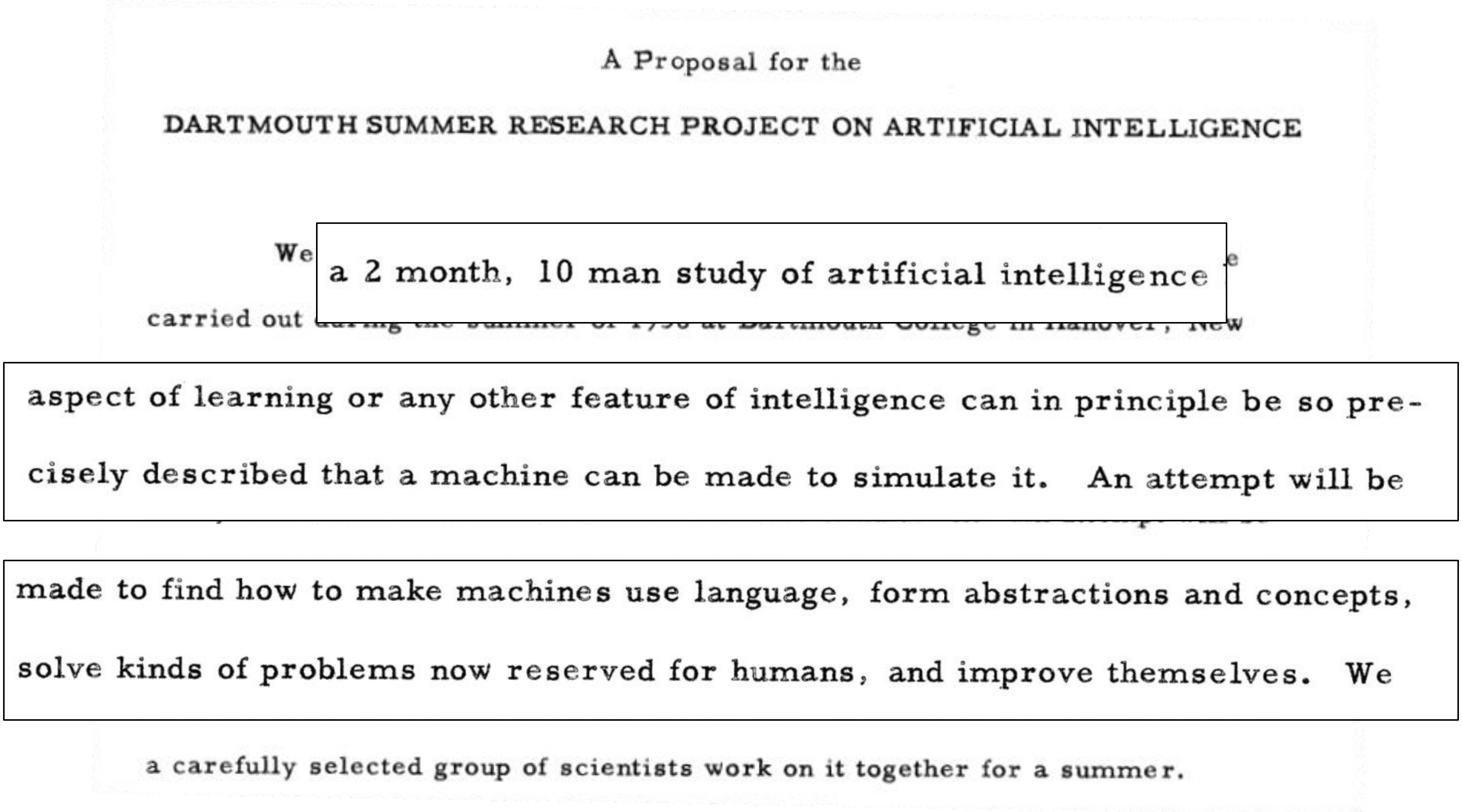
Fact: Ptolemy II‘s procession in 279 B.C. contained an automated statue of a god.

250 years ago

The Industrial Revolution

- population began to exhibit unprecedented sustained growth
- significant rise in the overall standard of living
- significant rise in education
- surge in technological capacity

1950s



A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

We

a 2 month, 10 man study of artificial intelligence

carried out

aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be

made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We

a carefully selected group of scientists work on it together for a summer.

John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (31 August 1955), p 1.

Defining AI

- John McCarthy went on to become one of the founders of the AI lab at Stanford University.

- His definition of AI was:

> the science and engineering of making intelligent machines.

- 1983 definition of AI by Elaine Rich:

> AI is the study of how to make computers do things at which, at the moment, people are better.

AI today

An (impressive) display of AI technology in a number of applications.

1997: Deep Blue – watch coverage here.

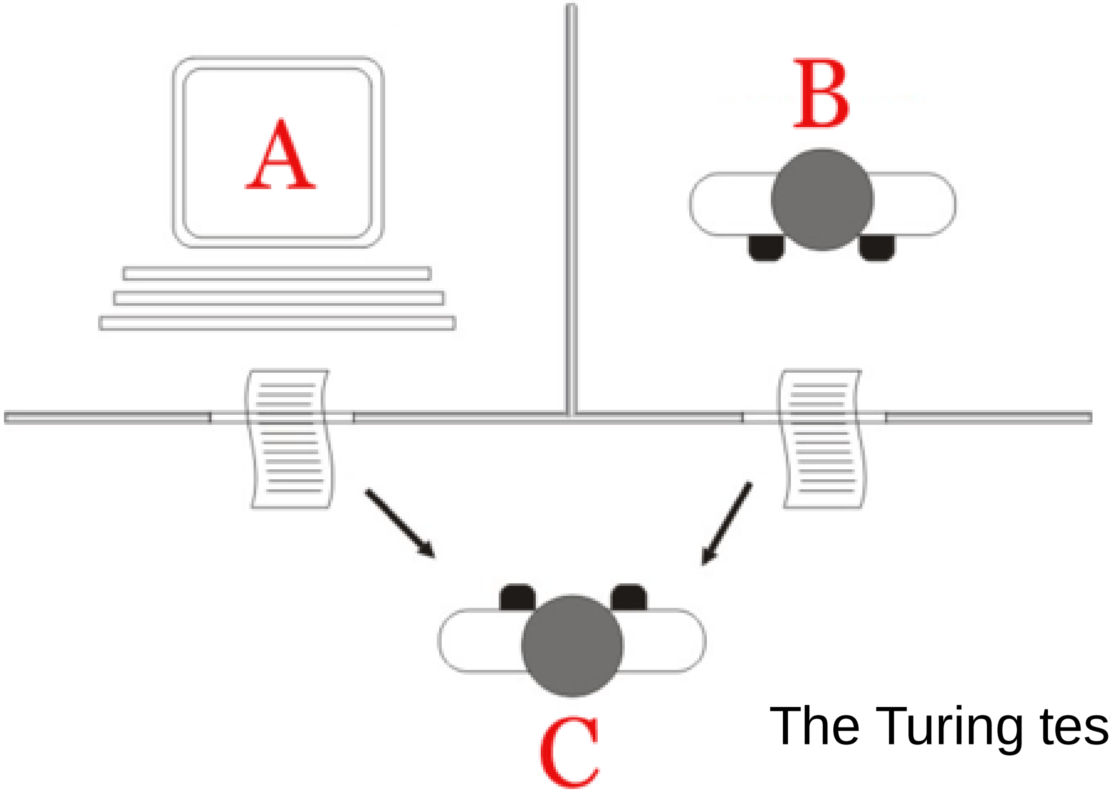
https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/

2011: Jeopardy – watch footage here.

2015: AlphaGo – watch footage here.

2018: Project Debater – watch footage here.

What does intelligent in AI mean?

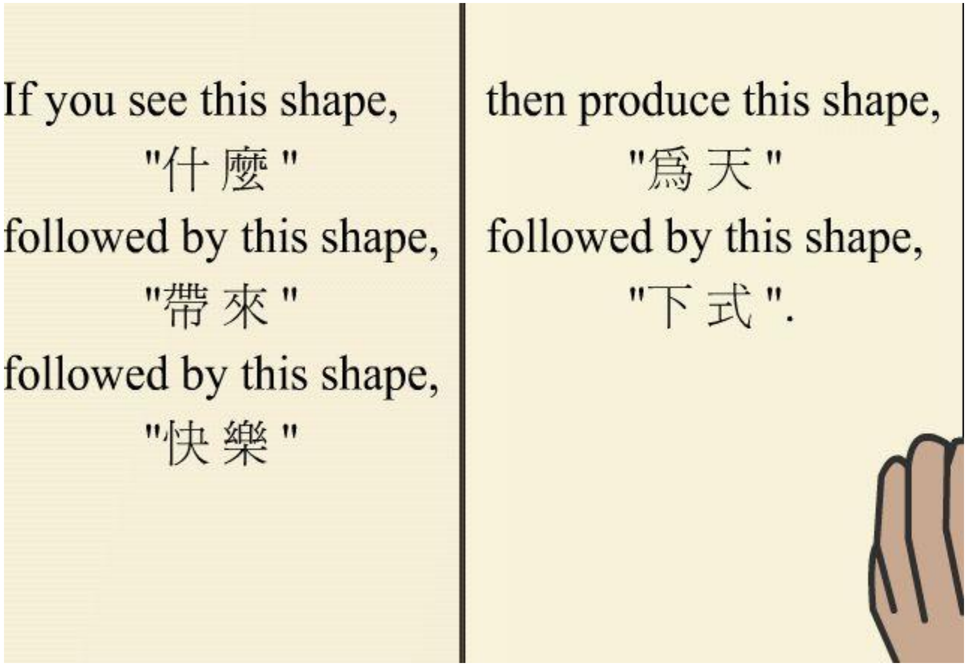


The Turing test, 1950.

But: John Searle’s Chinese Room Experiment (1980)

https://www.youtube.com/watch?v=TryOC83PH1g

But: John Searle’s Chinese Room Experiment



If you see this shape,
"什 麼 "
followed by this shape,
"帶 來 "
followed by this shape,
"快 樂 "

then produce this shape,
"爲 天 "
followed by this shape,
"下 式 ".

http://www.mind.ilstu.edu/curriculum/searle_chinese_room/searle_chinese_room.php

Strong AI: machine understands the task

Weak AI: machine follows a set of instructions according to which a task is performed.

- Shows that even if a machine is performing intelligent tasks, it does not necessarily actually “understand” that task in a meaningful way.
- Searle distinguished between “strong AI“ and “weak AI”
- The machines we are surrounded with so far are all examples of weak AI.

Strong versus weak AI

Strong versus weak AI, distinguishable by their goals:

“Strong” AI seeks to create artificial persons: machines that have all the mental powers we have, including phenomenal consciousness.

“Weak” AI, on the other hand, seeks to build information-processing machines that *appear* to have the full mental repertoire of human persons.

Searle’s Chinese Room experiment is designed to overthrow Strong AI.

The philosophy of AI

Explores artificial intelligence and its implications for knowledge and understanding of intelligence, ethics, consciousness, epistemology, and free will.

Prominent questions:

1. Can a machine act intelligently? Can it solve any problem that a person would solve by thinking?
2. Are human intelligence and machine intelligence the same? Is the human brain essentially a computer?
3. Can a machine have a mind, mental states, and consciousness in the same sense that a human being can? Can it feel how things are?

AI and consciousness

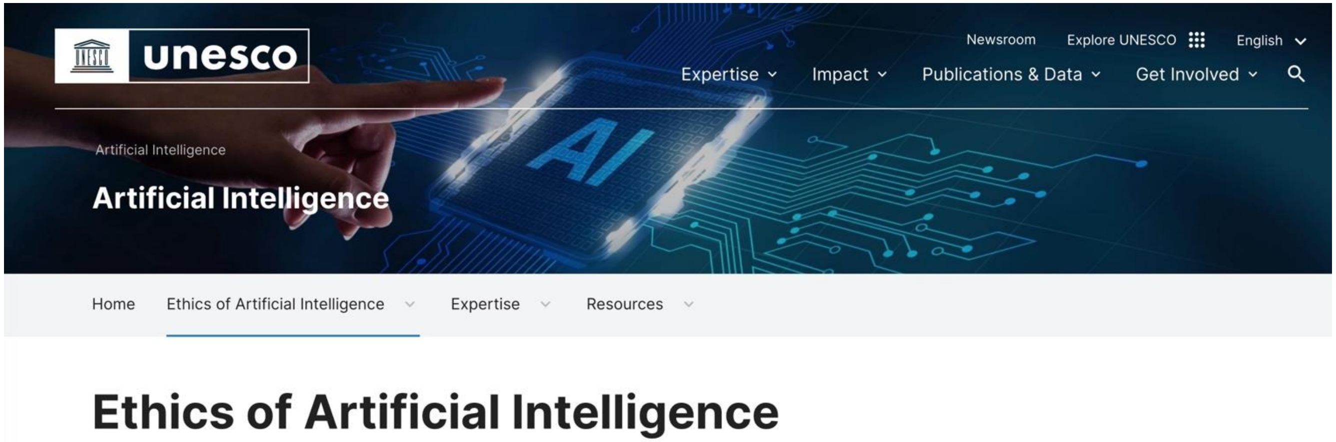
Taken from Hildt (2019).

Difficulty with artificial consciousness: What is consciousness at all? And how can subjectivity emerge from matter? (the “hard problem of consciousness” (Chalmers, 1996))

Human consciousness: available to us in the first-person perspective.

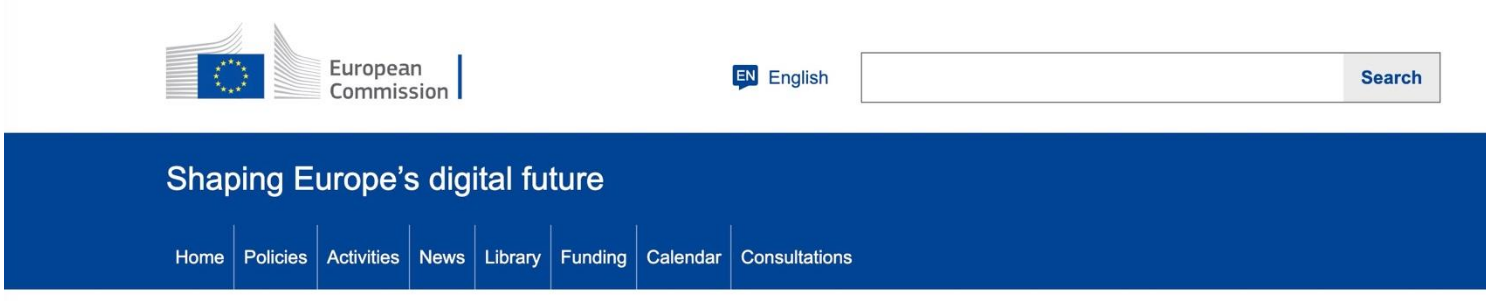
Artificial consciousness: only accessible for us in the third-person perspective (how do we know whether a machine has consciousness?)

AI and ethics



Artificial Intelligence

Ethics of Artificial Intelligence

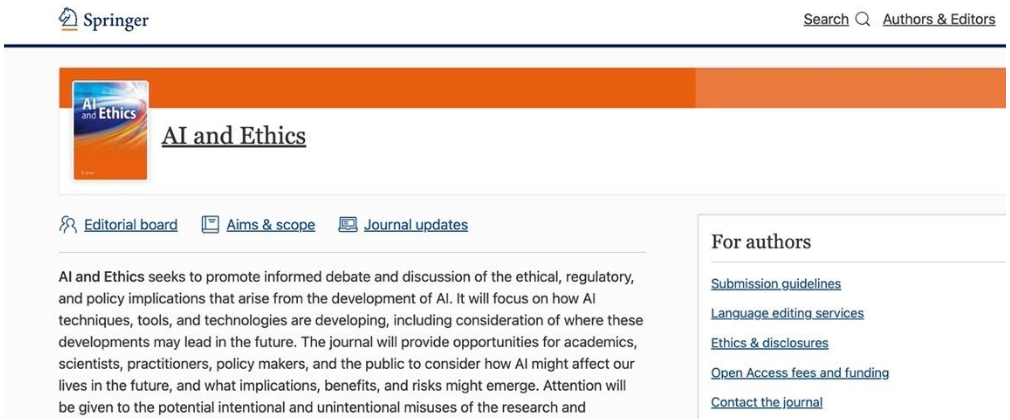


Shaping Europe's digital future

Home > Library > Ethics guidelines for trustworthy AI

REPORT / STUDY | Publication 08 April 2019

Ethics guidelines for trustworthy AI



AI and Ethics

AI and Ethics seeks to promote informed debate and discussion of the ethical, regulatory, and policy implications that arise from the development of AI. It will focus on how AI techniques, tools, and technologies are developing, including consideration of where these developments may lead in the future. The journal will provide opportunities for academics, scientists, practitioners, policy makers, and the public to consider how AI might affect our lives in the future, and what implications, benefits, and risks might emerge. Attention will be given to the potential intentional and unintentional misuses of the research and

For authors

- Submission guidelines
- Language editing services
- Ethics & disclosures
- Open Access fees and funding
- Contact the journal

AI and ethics

What are the challenges ahead?

- Evolution of the workforce
- Our physical and mental well-being
- The future of society

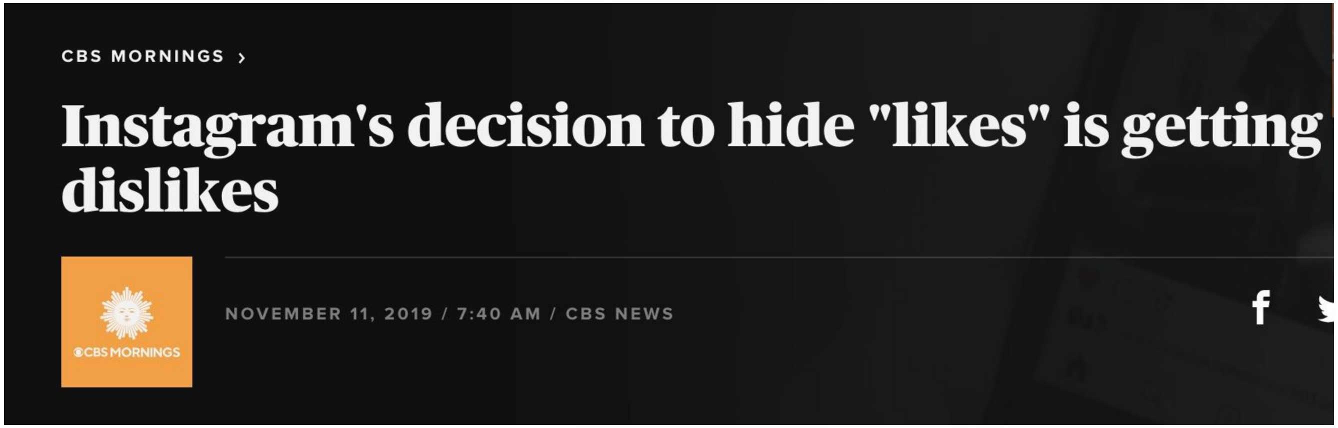
What are current efforts?

- Data protection and privacy regulation
- Law on using AI
- Call for a stop of further developing LLMs

AI and the workforce

- BBC article on situation in the UK -- 7 million jobs could be replaced by AI, but 7.2 million could be created.
- AI takes over repetitive or dangerous tasks, let humans do tasks requiring creativity and empathy.
- Enhance monitoring and diagnosing capabilities in medicine, cheaper healthcare (McKinsey report from 2013)
- Uncover criminal activity and solve crimes.

AI and our health



CBS MORNINGS ›

Instagram's decision to hide "likes" is getting dislikes

NOVEMBER 11, 2019 / 7:40 AM / CBS NEWS

Humankind has never been as connected as it is now.

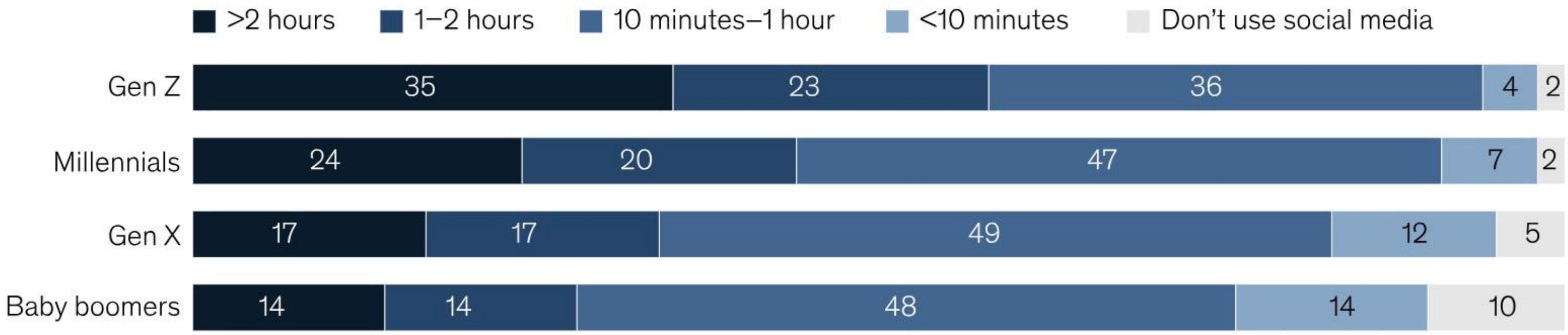
McKinsey report from April 2023:

- More than 50% of people across age cohorts cite self-expression and social connectivity as positives of social media.
- Complex relationship between mental health and social media: there is correlation, but hard to identify causation.

AI and our health

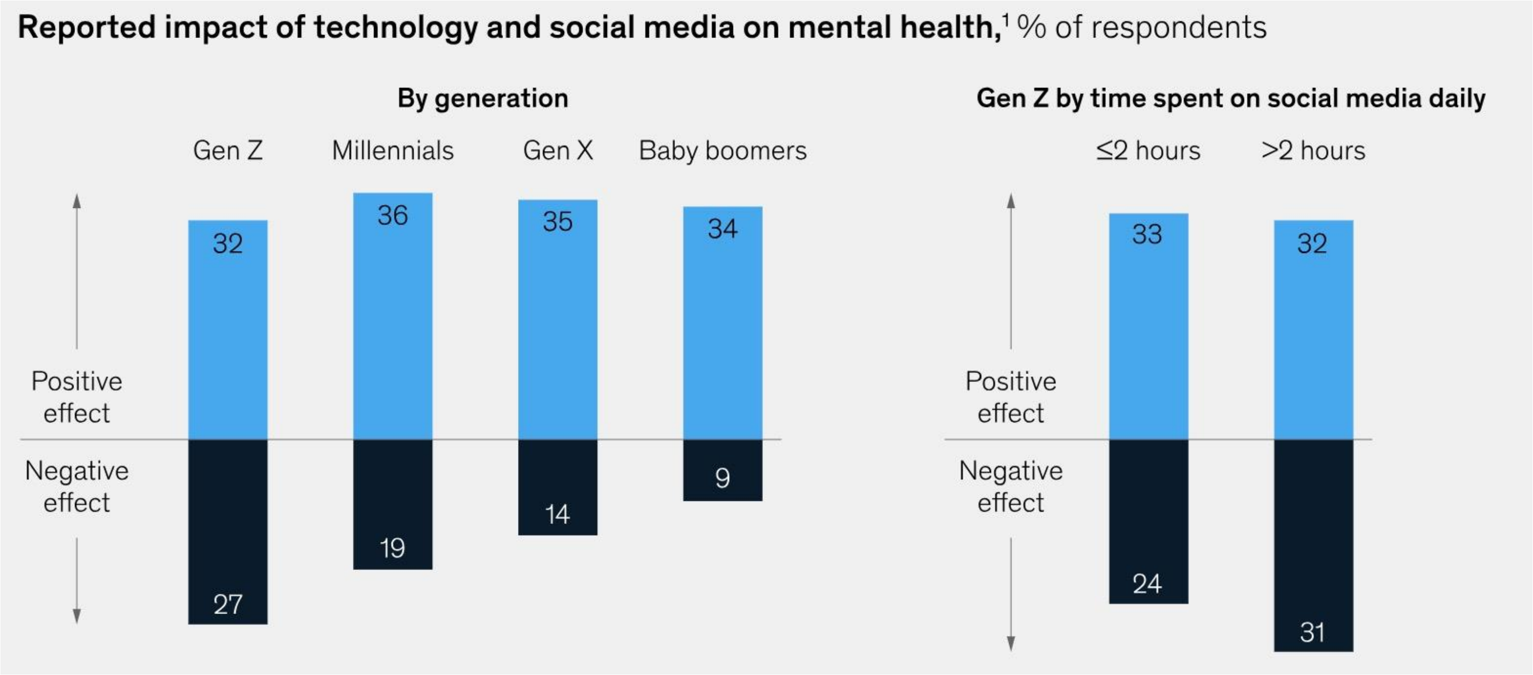
Almost everyone is using social media, but in different ways.

Time spent on social media daily,[1] % of respondents (n = 41,960)



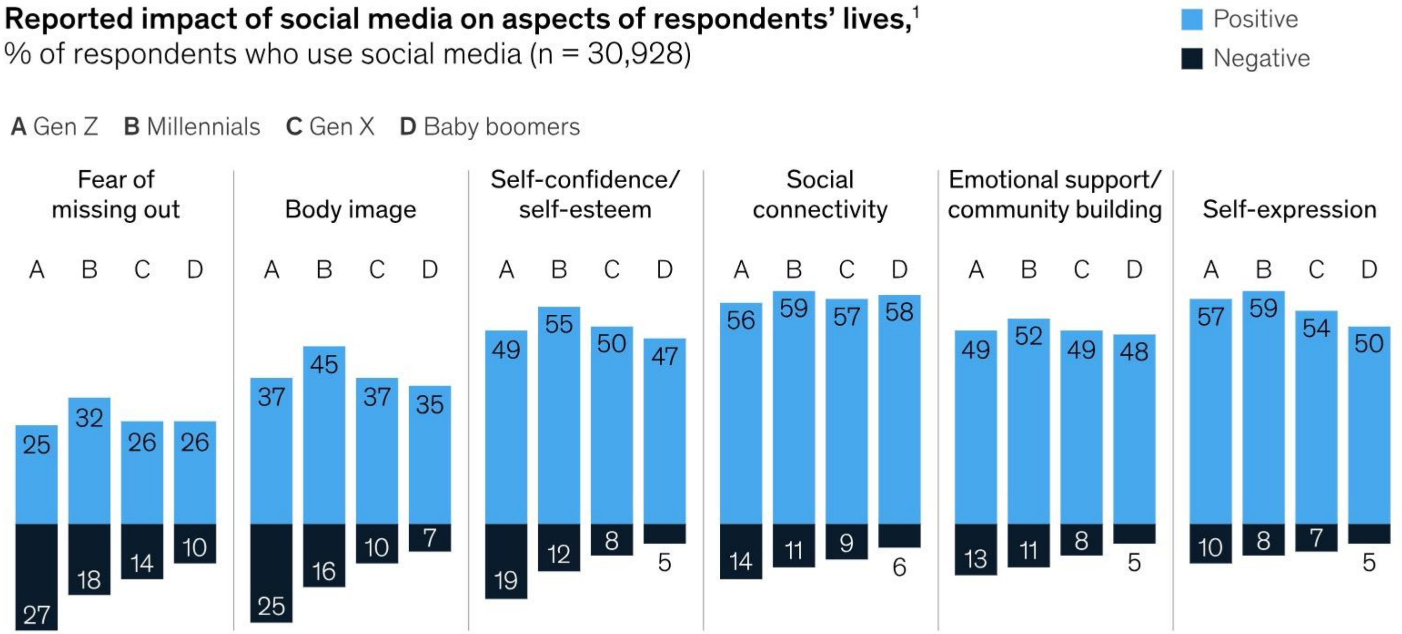
AI and our health

While social media and tech have a consistent positive impact across all age cohorts, the negative impact increases substantially for young ages.



AI and our health

Respondents’ assessment of the impact of social media ranges substantially depending on the dimension.



Reported impact of social media on aspects of respondents’ lives,[1]
% of respondents who use social media (n = 30,928)

AI and our health

Mental health

Association of Facebook Use With Compromised Well-Being: A Longitudinal Study, Holly B. Shakya, Nicholas A. Christakis , 2017, *American Journal of Epidemiology*, Volume 185, Issue 3, 1 February 2017, Pages 203–211.

5,208 subjects: overall, regular use of Facebook had a negative impact on an individual’s wellbeing

AI and healthcare

European Parliament

Artificial intelligence in healthcare

Applications, risks, and ethical and societal impacts

What are the risks of using AI in in medicine and healthcare?
https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf

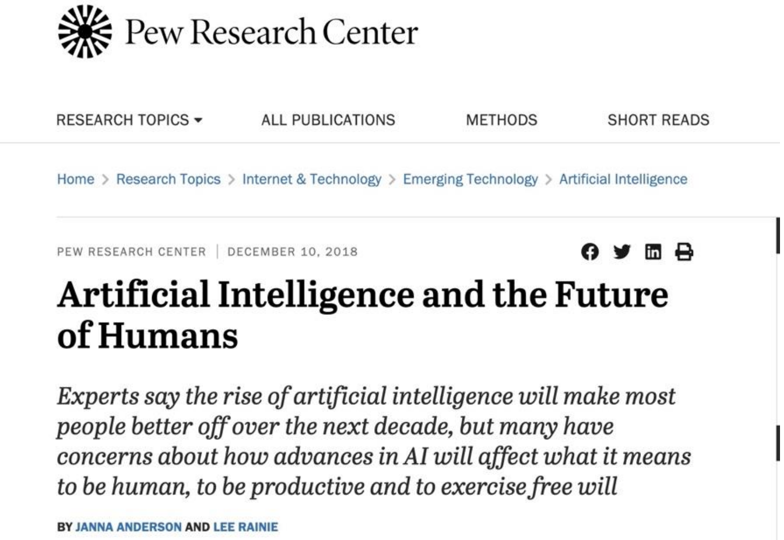
What are its benefits? Search online.

AI and the future of society

Wide variety of viewpoints how the future of humankind and society will develop when AI becomes a participant.

Basic questions:

- In which areas can humanity benefit?
- What are the dangers?



Pew Research Center

RESEARCH TOPICS ▾ ALL PUBLICATIONS METHODS SHORT READS

Home > Research Topics > Internet & Technology > Emerging Technology > Artificial Intelligence

PEW RESEARCH CENTER | DECEMBER 10, 2018

Artificial Intelligence and the Future of Humans

Experts say the rise of artificial intelligence will make most people better off over the next decade, but many have concerns about how advances in AI will affect what it means to be human, to be productive and to exercise free will

BY JANNA ANDERSON AND LEE RAINIE

https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/

Funders pick up on it: VolkswagenStiftung's AI and the future of society

AI and the future of society

Major concerns (PEW paper):

- **Human agency:** Humans experience a loss of control over their lives (e.g., privacy, decision-making)
- **Data abuse:** Data use and surveillance in complex systems is designed for profit or for exercising power
- **Job loss:** The AI takeover of jobs will widen economic divides, leading to social upheaval
- **Dependence lock-in:** Reduction of individuals’ cognitive, social and survival skills
- **Mayhem:** Autonomous weapons, cybercrime and weaponized information.

AI and the future of society

Suggested solutions:

- **Global good is #1:** Improve human collaboration across borders and stakeholder groups (“make people around the world come to a common understanding and agreement”)
- **Values-based system:** Develop policies to assure AI will be directed at ‘humanness’ and common good
- **Prioritize people:** Alter economic and political systems to better help humans ‘race with the robots’

The AI Index Report 2024

Human-Centered Artificial Intelligence, Stanford University

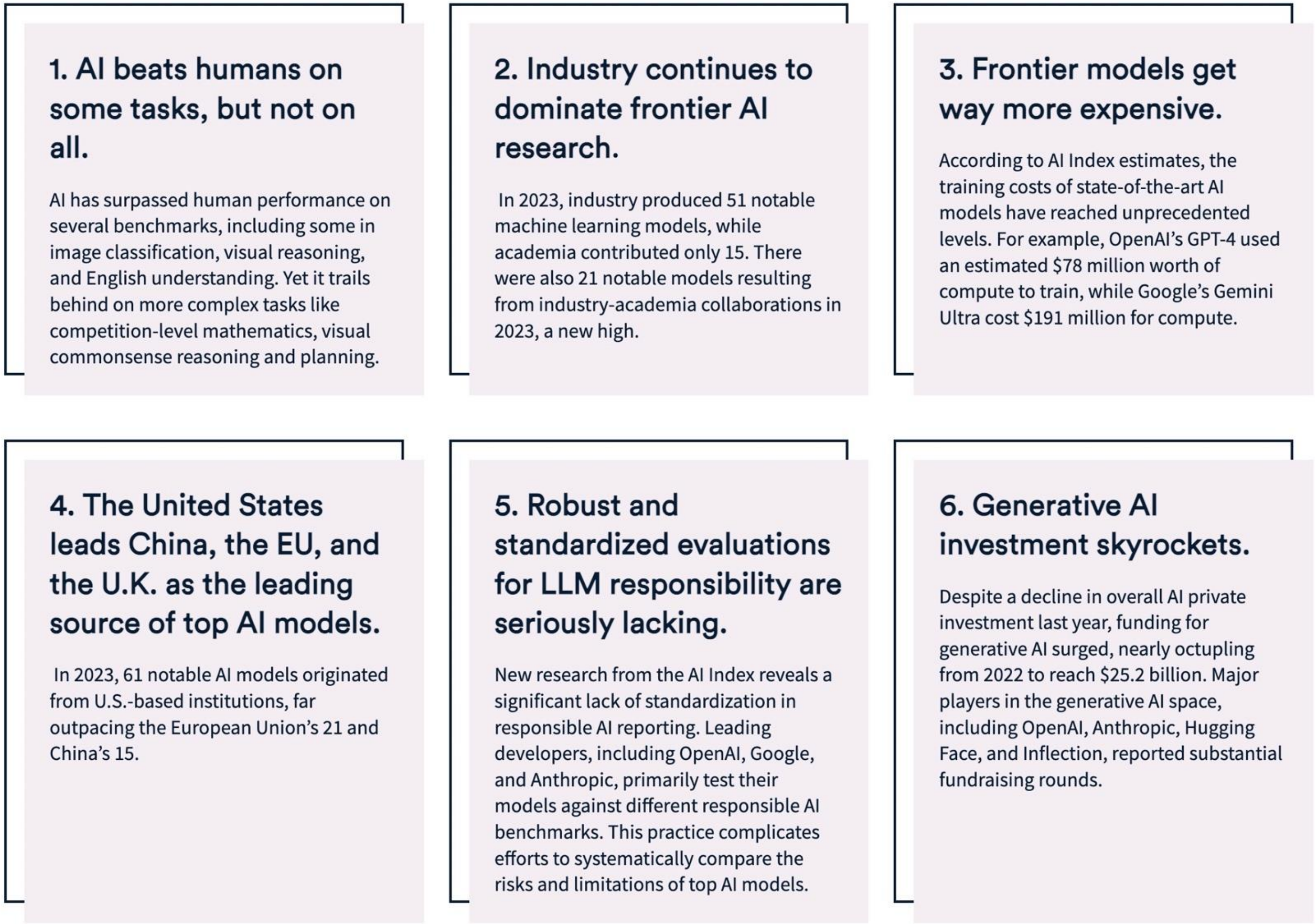
Yearly report on how the field of AI has involved.

- unbiased, rigorously vetted, broadly sourced data basis
- allow the public and policy makers to develop a more thorough and nuanced understanding of AI

At the forefront this year: Generative AI

taken from https://aiindex.stanford.edu/report/

The AI Index Report - Top takeaways in 2024



1. AI beats humans on some tasks, but not on all.

AI has surpassed human performance on several benchmarks, including some in image classification, visual reasoning, and English understanding. Yet it trails behind on more complex tasks like competition-level mathematics, visual commonsense reasoning and planning.

2. Industry continues to dominate frontier AI research.

In 2023, industry produced 51 notable machine learning models, while academia contributed only 15. There were also 21 notable models resulting from industry-academia collaborations in 2023, a new high.

3. Frontier models get way more expensive.

According to AI Index estimates, the training costs of state-of-the-art AI models have reached unprecedented levels. For example, OpenAI's GPT-4 used an estimated $78 million worth of compute to train, while Google's Gemini Ultra cost $191 million for compute.

4. The United States leads China, the EU, and the U.K. as the leading source of top AI models.

In 2023, 61 notable AI models originated from U.S.-based institutions, far outpacing the European Union's 21 and China's 15.

5. Robust and standardized evaluations for LLM responsibility are seriously lacking.

New research from the AI Index reveals a significant lack of standardization in responsible AI reporting. Leading developers, including OpenAI, Google, and Anthropic, primarily test their models against different responsible AI benchmarks. This practice complicates efforts to systematically compare the risks and limitations of top AI models.

6. Generative AI investment skyrockets.

Despite a decline in overall AI private investment last year, funding for generative AI surged, nearly octupling from 2022 to reach $25.2 billion. Major players in the generative AI space, including OpenAI, Anthropic, Hugging Face, and Inflection, reported substantial fundraising rounds.

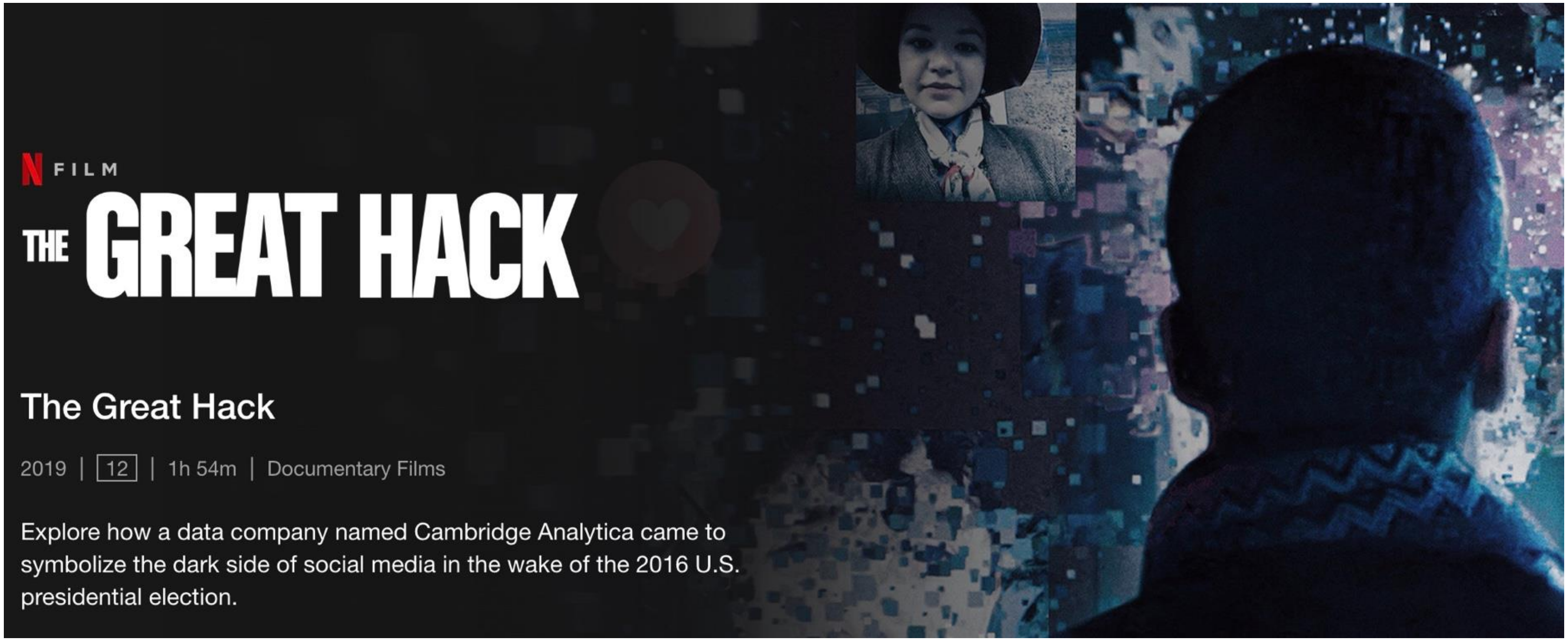
taken from https://aiindex.stanford.edu/report/

Generative AI

What are the risks and benefits of Generative AI? Search online.

Regulating AI

‘The Great Hack’ (Netflix documentation – trailer)



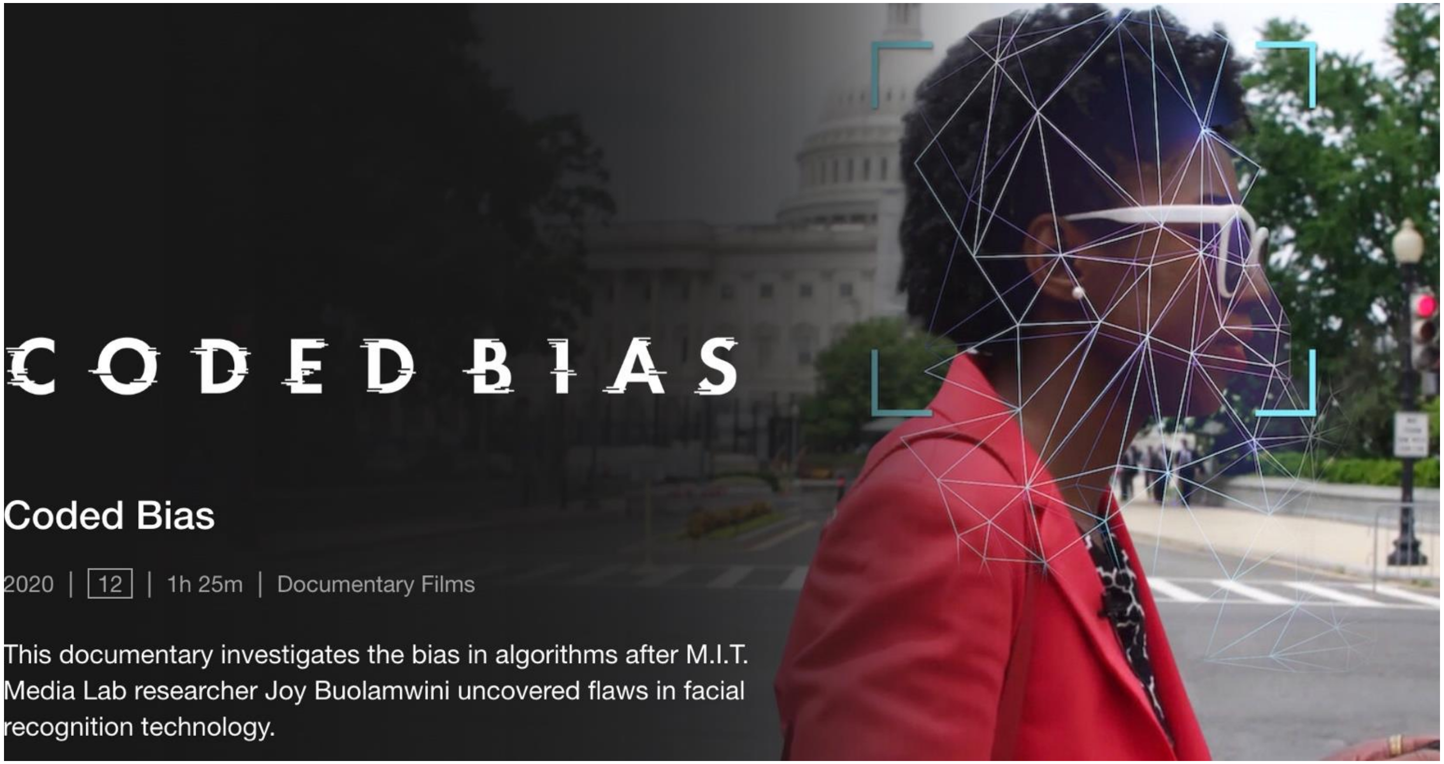
The Great Hack

2019 | 12 | 1h 54m | Documentary Films

Explore how a data company named Cambridge Analytica came to symbolize the dark side of social media in the wake of the 2016 U.S. presidential election.

Regulating AI

‘Coded bias’ (Netflix documentation – trailer)



CODED BIAS

Coded Bias

2020 | 12 | 1h 25m | Documentary Films

This documentary investigates the bias in algorithms after M.I.T. Media Lab researcher Joy Buolamwini uncovered flaws in facial recognition technology.

Regulating AI

Regulators must act on the risks presented by new technology.

→ We need laws that regulate AI applications so that we can all benefit, but nobody gets “hurt”.

Europe is spear-heading these efforts.

First step: Data protection

It all started in 1995 with the European Data Protection Directive.

- In 1994, the first banner add appeared online.
- In 2000, most financial institutions offered online banking.
- In 2006, Facebook opened to the public.
- In 2011, a Google user sued the company for scanning her emails.
- Two months later: The EU needs “a comprehensive approach on personal data protection” and work begins to update the 1995 directive.

First step: Data protection

The General Data Protection Regulation (GDPR) is the toughest privacy and security law in the world.

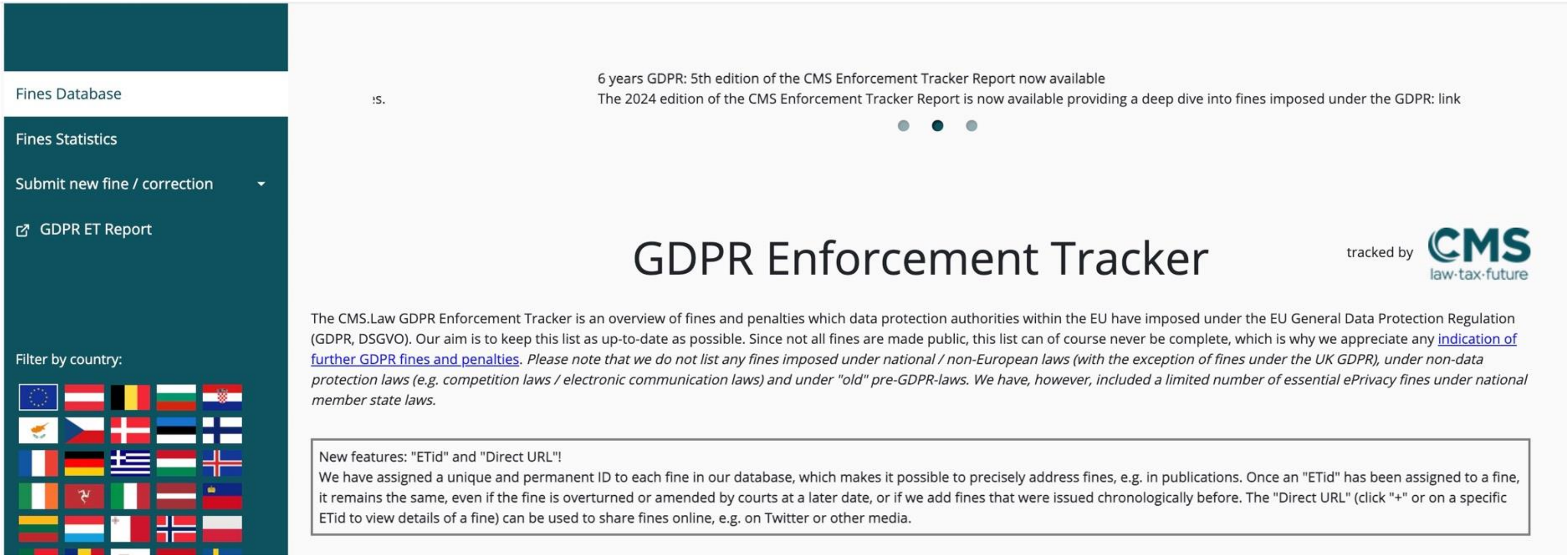
- Drafted and passed by the European Union.

- Put into effect on May 25, 2018.

- Imposes obligations onto organizations anywhere in the world if they target or collect data related to people in the EU.

- European Convention on Human Rights (1950): ”Everyone has the right to respect for his private and family life, his home and his correspondence.”

GDPR breaches: Facebook



6 years GDPR: 5th edition of the CMS Enforcement Tracker Report now available
The 2024 edition of the CMS Enforcement Tracker Report is now available providing a deep dive into fines imposed under the GDPR: link

GDPR Enforcement Tracker

The CMS.Law GDPR Enforcement Tracker is an overview of fines and penalties which data protection authorities within the EU have imposed under the EU General Data Protection Regulation (GDPR, DSGVO). Our aim is to keep this list as up-to-date as possible. Since not all fines are made public, this list can of course never be complete, which is why we appreciate any indication of further GDPR fines and penalties. *Please note that we do not list any fines imposed under national / non-European laws (with the exception of fines under the UK GDPR), under non-data protection laws (e.g. competition laws / electronic communication laws) and under "old" pre-GDPR-laws. We have, however, included a limited number of essential ePrivacy fines under national member state laws.*

New features: "ETid" and "Direct URL"!
We have assigned a unique and permanent ID to each fine in our database, which makes it possible to precisely address fines, e.g. in publications. Once an "ETid" has been assigned to a fine, it remains the same, even if the fine is overturned or amended by courts at a later date, or if we add fines that were issued chronologically before. The "Direct URL" (click "+" or on a specific ETid to view details of a fine) can be used to share fines online, e.g. on Twitter or other media.

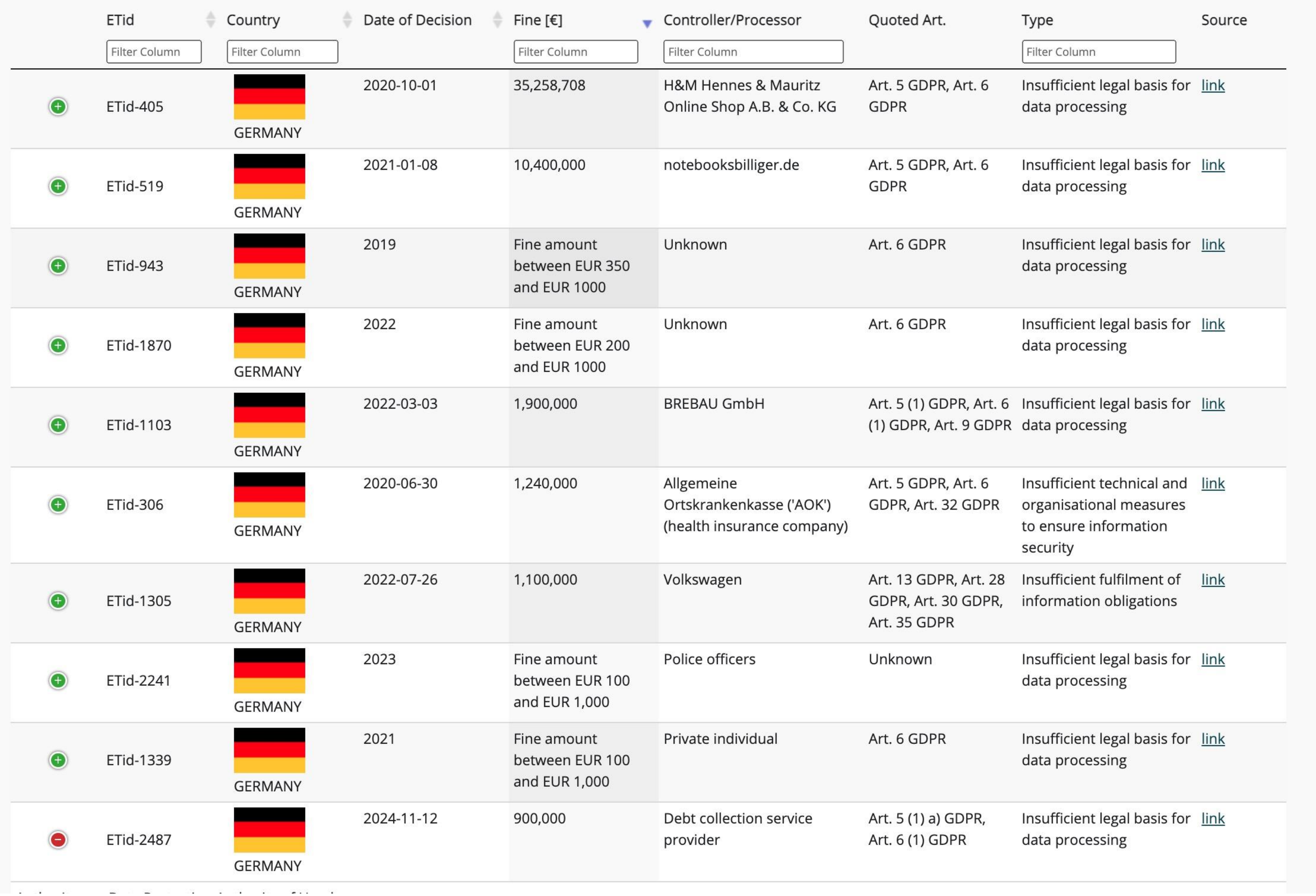
https://www.enforcementtracker.com/

GDPR breaches

- Unlawful transfer of personal data to the US
  - The Irish DPA against Meta Platforms Ireland Limited in the amount of EUR 1.2 billion
  - The Dutch DPA agency against Uber in the amount of EUR 290 million
- Insufficient technical measure to secure personal information
  - The Irish DPA against Meta in the amount of EUR 250 million

…

GDPR breaches: Google



	ETid	Country	Date of Decision	Fine [€]	Controller/Processor	Quoted Art.	Type	Source
⊕	ETid-405	GERMANY	2020-10-01	35,258,708	H&M Hennes & Mauritz Online Shop A.B. & Co. KG	Art. 5 GDPR, Art. 6 GDPR	Insufficient legal basis for data processing	link
⊕	ETid-519	GERMANY	2021-01-08	10,400,000	notebooksbilliger.de	Art. 5 GDPR, Art. 6 GDPR	Insufficient legal basis for data processing	link
⊕	ETid-943	GERMANY	2019	Fine amount between EUR 350 and EUR 1000	Unknown	Art. 6 GDPR	Insufficient legal basis for data processing	link
⊕	ETid-1870	GERMANY	2022	Fine amount between EUR 200 and EUR 1000	Unknown	Art. 6 GDPR	Insufficient legal basis for data processing	link
⊕	ETid-1103	GERMANY	2022-03-03	1,900,000	BREBAU GmbH	Art. 5 (1) GDPR, Art. 6 (1) GDPR, Art. 9 GDPR	Insufficient legal basis for data processing	link
⊕	ETid-306	GERMANY	2020-06-30	1,240,000	Allgemeine Ortskrankenkasse ('AOK') (health insurance company)	Art. 5 GDPR, Art. 6 GDPR, Art. 32 GDPR	Insufficient technical and organisational measures to ensure information security	link
⊕	ETid-1305	GERMANY	2022-07-26	1,100,000	Volkswagen	Art. 13 GDPR, Art. 28 GDPR, Art. 30 GDPR, Art. 35 GDPR	Insufficient fulfilment of information obligations	link
⊕	ETid-2241	GERMANY	2023	Fine amount between EUR 100 and EUR 1,000	Police officers	Unknown	Insufficient legal basis for data processing	link
⊕	ETid-1339	GERMANY	2021	Fine amount between EUR 100 and EUR 1,000	Private individual	Art. 6 GDPR	Insufficient legal basis for data processing	link
⊖	ETid-2487	GERMANY	2024-11-12	900,000	Debt collection service provider	Art. 5 (1) a) GDPR, Art. 6 (1) GDPR	Insufficient legal basis for data processing	link

Second step: Regulating actual AI systems

Legal and regulatory frameworks typically operate around a clear sense of who is acting, what their mindset was at the time of action and where the action takes place.

Problem when applied to AI!

Examples:

- *Who is liable for accidents if a car is driverless?*

- *What recourse does an individual have if it is refused from insurance based on an automatic analysis of their social media accounts?*

Second step: Regulating actual AI systems

The European Commission (EC): Shaping Europe’s digital future

https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

The EU’s approach to AI rests on excellence and trust.

Aim: Boost research and industrial capacity and ensure fundamental rights.

“Europe as the global hub for trustworthy AI.“

Second step: Regulating actual AI systems

The EU AI Act

Put in place 1 August 2024

Ethical guidelines and a regulatory framework for the development, deployment, and use of Artificial Intelligence (AI) systems in the European Union

- Different rules for different risk levels
- Transparency requirement
- Supporting innovation

https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Second step: Regulating actual AI systems

Ethics guidelines for trustworthy AI by the EU

Put forth in 2019 by the High-Level Expert Group on AI.

Trustworthy AI should be:

- lawful - respecting all applicable laws and regulations
- ethical - respecting ethical principles and values
- robust - both from a technical perspective while taking into account its social environment

The piloting phase ended in December 2019.

Second step: Regulating actual AI systems

Result: Translation of the ethics guidelines into an

“accessible and dynamic (self-assessment) checklist. The checklist can be used by developers and deployers of AI who want to implement the key requirements in practice.” (https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai)

→ What are the important (subtext) words here?

Second step: Regulating actual AI systems

Reaction by Thomas Metzinger, Philosoph, member of the EU’s expert panel for developing the ethics guidelines (here in German).

The guidelines are:

- short-sighted
- deliberately vague
- do not take long-term risks into consideration

Red lines were deleted or watered down in the final report (for example, autonomous lethal weapons and social scoring systems).

Big problem: No regulatory oversight to support implementation

Second step: Regulating actual AI systems

What about Generative AI?

Generative AI, like ChatGPT, will not be classified as high-risk, but will have to comply with transparency requirements and EU copyright law:

- Disclosing that the content was generated by AI
- Designing the model to prevent it from generating illegal content
- Publishing summaries of copyrighted data used for training

https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Google on regulation

Google’s recommendations for regulating AI.

CEO Sundar Pichai:

- “AI is too important not to regulate, the only question is how.”
- “Industry cannot do it alone but needs governments to guide the process.”

Today

Part 1: Artificial Intelligence

- A brief history
- Strong versus weak AI
- AI, ethics and regulation

Part 2: Superintelligence

Superintelligence

Superintelligence ~ Digital Superintelligence ~ Artificial General Intelligence ~ Singularity

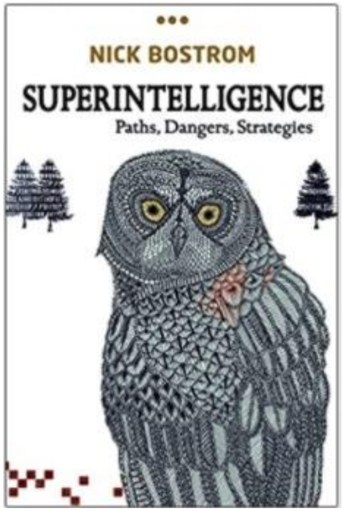
Research on superintelligence asks the questions:

1. What happens when machines surpass humans in general intelligence?

2. Will artificial agents save or destroy us?

“Machine intelligence is the last invention that humanity will ever need to make.”

Superintelligence



Nick Bostrom, Director, Future of Humanity Institute, University of Oxford, UK.

In 2009 and 2015, he was included in *Foreign Policy*'s Top 100 Global Thinkers list.

His book: Superintelligence – Paths, Dangers, Strategies, Oxford University Press, 2014: Understand the challenge posed by the prospect of superintelligence and how we respond best.

Superintelligence as "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest”.

Superintelligence

If machine brains surpassed human brains in general intelligence, then this new superintelligence could become very powerful – possibly beyond our control.

Compare with the fate of the gorilla: their survival depends on the humans, more than on themselves.

One advantage of humans: we make the first move!

Question: Will it be possible to construct an AI with initial conditions so that we survive an intelligence explosion?

Bostrom’s TED Talk "*What happens when our computers get smarter than we are?“* here:
https://www.youtube.com/watch?v=MnT1xgZgkpk

The timescale

Will singularity ever happen? According to most AI experts, yes.
When will the singularity happen? Before the end of the century

But the views are divided.

See the survey of 21 AI experts at the ‘Artificial General Intelligence’ Conference in 2009

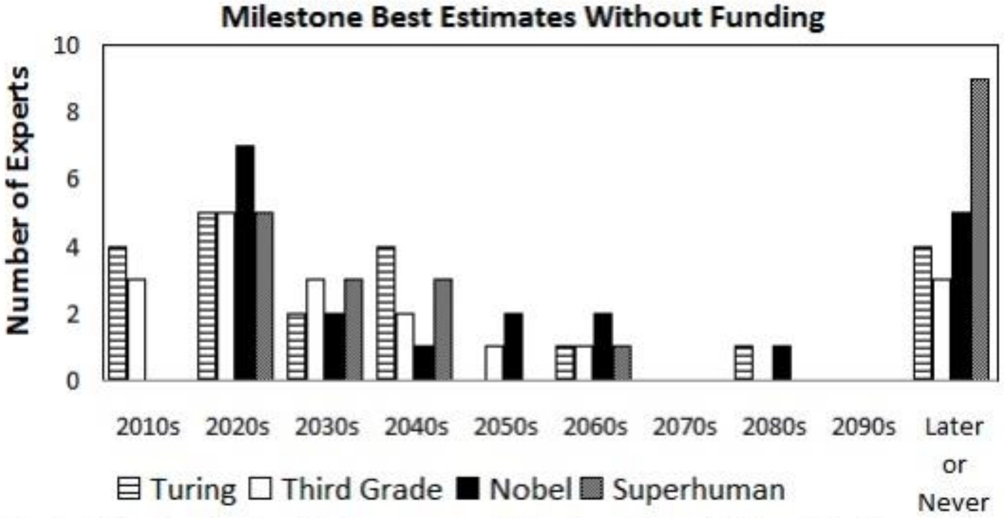


Fig. 2: Milestone best estimate guesses without massive additional funding. Estimates are for when AI would achieve four milestones: the Turing Test (horizontal lines), third grade (white), Nobel-quality work (black), and superhuman capability (grey).

The timescale

Another survey, conducted in 2012/2013 by Nick Bostrom and Vincent C. Muller, the president of the European Association for Cognitive Systems.

550 participants answered the question: “When is AGI likely to happen?”

The answers are distributed as:

- 10% of participants think that AGI is likely to happen by 2022
- For 2040, the share is 50%
- 90% of participants think that AGI is likely to happen by 2075.

The timescale

In 2017, 352 AI experts who published at the 2015 NIPS (Neural Information Processing Systems) and ICML (International Conference on Machine Learning) conferences were surveyed.

Results:

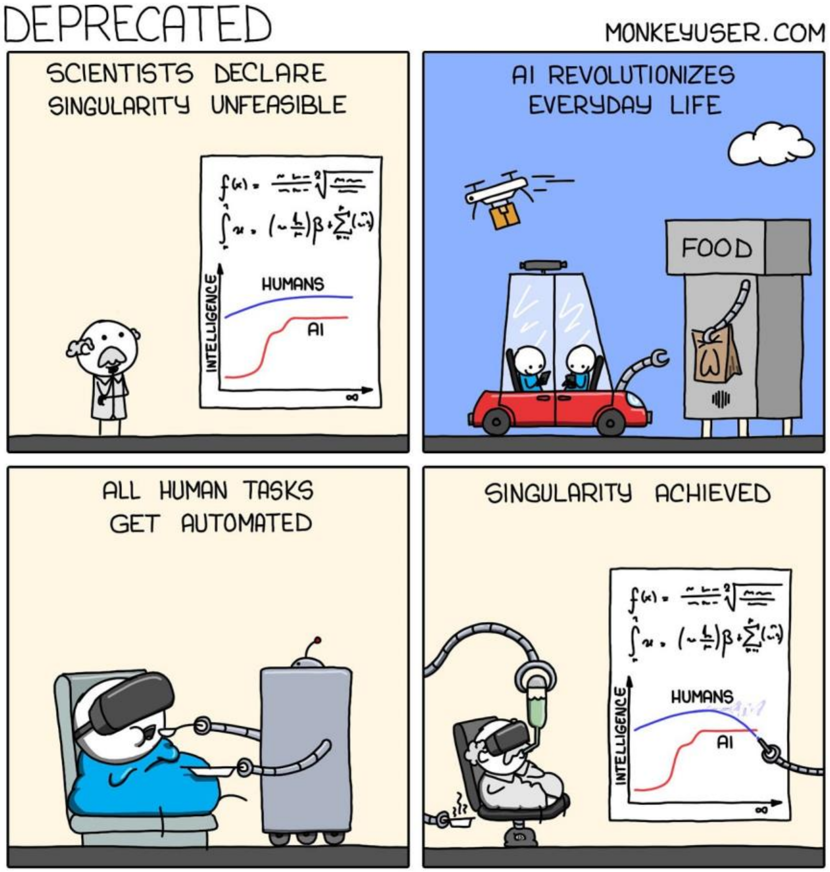
- 50% chance that AGI will occur until 2060.
- Significant difference of opinion based on geography:
  - Asian respondents expect AGI in 30 years
  - North Americans expect it in 74 years.
- Some significant job functions that are expected to be automated until 2030 are: Call center reps, truck driving, retail sales.

Cluster of Excellence
The Politics of Inequality

Steinbeis Transfer Center
Linguistic Data Analysis

UNIVERSITÄT PASSAU

Thank you.
Questions?
Comments?



DEPRECATED
MONKEYUSER.COM
SCIENTISTS DECLARE SINGULARITY UNFEASIBLE
INTELLIGENCE
HUMANS
AI
∞
AI REVOLUTIONIZES EVERYDAY LIFE
FOOD
ALL HUMAN TASKS GET AUTOMATED
SINGULARITY ACHIEVED
INTELLIGENCE
HUMANS
AI
∞

Annette Hautli-Janisz, Prof. Dr.
cornlp-teaching@uni-passau.de