Foundations of Ethical Artificial Intelligence

Foundations of Ethical Artificial Intelligence

Computational Perspective

Roberta Calegari
roberta.calegari@unibo.it

Alma Mater Studiorum – Università di Bologna

10 October 2024

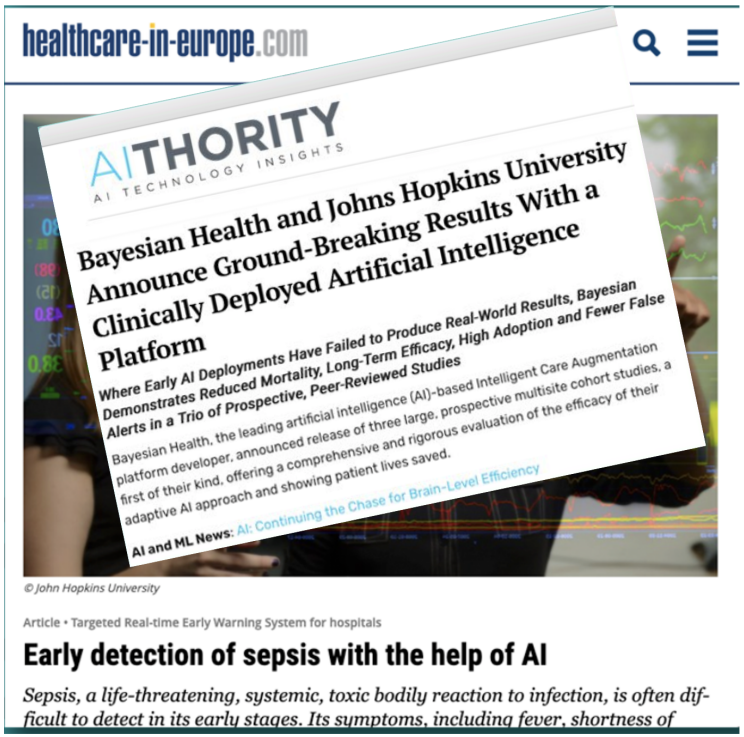
Outline

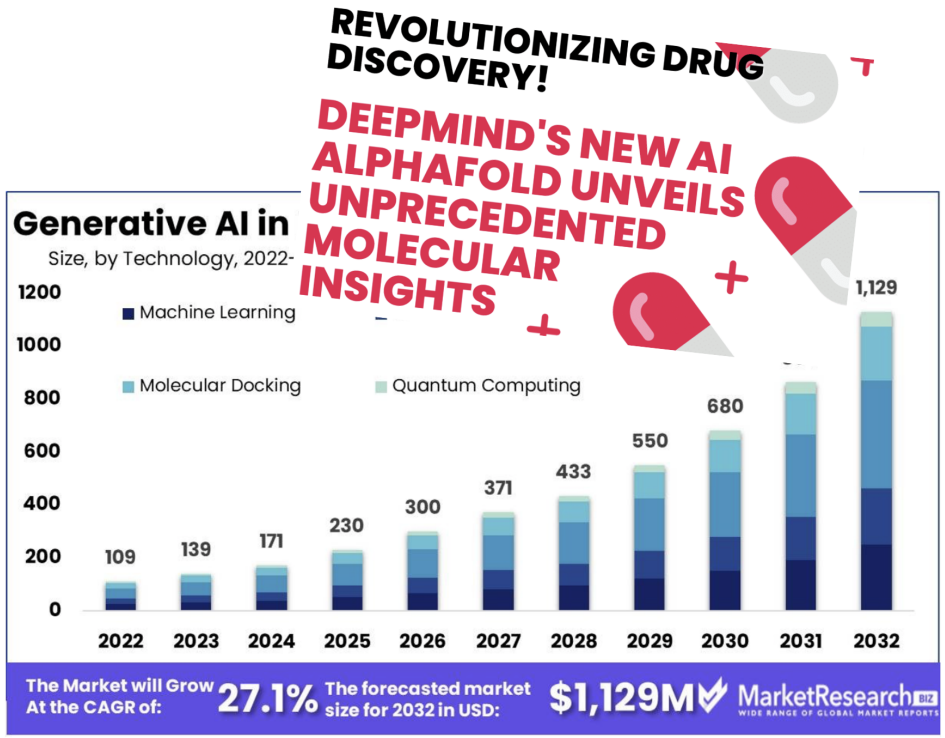
1. Ethics of AI: why?
2. Ethics of AI: initiatives
3. Ethics of AI: EGTAI
4. EGTAI Definitions & Examples
5. AI for Social Good
6. Summing up

Next in Line...

1. Ethics of AI: why?
2. Ethics of AI: initiatives
3. Ethics of AI: EGTAI
4. EGTAI Definitions & Examples
5. AI for Social Good
6. Summing up

AI media I


healthcare-in-europe.com
AITHORITY
AI TECHNOLOGY INSIGHTS
Bayesian Health and Johns Hopkins University Announce Ground-Breaking Results With a Clinically Deployed Artificial Intelligence Platform
Where Early AI Deployments Have Failed to Produce Real-World Results, Bayesian Demonstrates Reduced Mortality, Long-Term Efficacy, High Adoption and Fewer False Alerts in a Trio of Prospective, Peer-Reviewed Studies
Bayesian Health, the leading artificial intelligence (AI)-based Intelligent Care Augmentation platform developer, announced release of three large, prospective multisite cohort studies, a first of their kind, offering a comprehensive and rigorous evaluation of the efficacy of their adaptive AI approach and showing patient lives saved.
AI and ML News: AI: Continuing the Chase for Brain-Level Efficiency
© John Hopkins University
Article • Targeted Real-time Early Warning System for hospitals
Early detection of sepsis with the help of AI
Sepsis, a life-threatening, systemic, toxic bodily reaction to infection, is often difficult to detect in its early stages. Its symptoms, including fever, shortness of

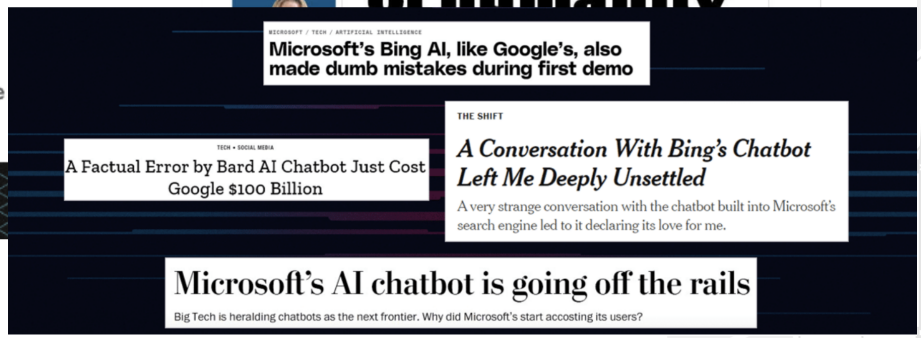

REVOLUTIONIZING DRUG DISCOVERY!
DEEPMIND'S NEW AI ALPHAFOLD UNVEILS UNPRECEDENTED MOLECULAR INSIGHTS
Generative AI in
Size, by Technology, 2022-
Machine Learning
Molecular Docking
Quantum Computing
1200
1000
800
600
400
200
0
109
139
171
230
300
371
433
550
680
1,129
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
The Market will Grow At the CAGR of:
27.1%
The forecasted market size for 2032 in USD:
$1,129M
MarketResearch.biz
WIDE RANGE OF GLOBAL MARKET REPORTS

AI media II


WILL A ROBOT STEAL YOUR JOB?
KILLER COMPUTERS
Bill Gates warns 'dangerous AI' poses a threat 'like nuclear weapons'
AI WARNING:
Robots will destroy a HUGE number of jobs, claims expert
AI could be used to TAKE OVER the WORLD through 'evil' fake news and hijacking cars
HIGH QUALITY
UPDATES
COVID
VIDEO


AI pioneers fear extinction
Our creations are as great a threat to humanity as nuclear war or pandemics, say hundreds of experts in call to regulate tech
SEE PAGES 18-19
A.I. 'COULD WIPE OUT HUMANITY'
Threat 'as bad as nuclear war' MUST be tackled, say tech bosses
WEDNESDAY
31 MAY 2023
FAMILY
I refuse to be called grandma - I hate it
AI creators fear the extinction of humanity


01-30-20
AI can now design cities. Should we let it?
We used to worry about AI turning our world into gray goo. But maybe afraid of it bringing in a Starbucks and a Lululemon.


Microsoft's Bing AI, like Google's, also made dumb mistakes during first demo
A Factual Error by Bard AI Chatbot Just Cost Google $100 Billion
THE SHIFT
A Conversation With Bing's Chatbot Left Me Deeply Unsettled
A very strange conversation with the chatbot built into Microsoft's search engine led to it declaring its love for me.
Microsoft's AI chatbot is going off the rails
Big Tech is heralding chatbots as the next frontier. Why did Microsoft's start accosting its users?

Science Finctiion Narration I

AI facts I

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

8 October 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Physics 2024 to

John J. Hopfield
Princeton University, NJ, USA

Geoffrey E. Hinton
University of Toronto, Canada

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

They trained artificial neural networks using physics

This year's two Nobel Laureates in Physics have used tools from physics to develop methods that are the foundation of today's powerful machine learning. John Hopfield created an associative memory that can store and reconstruct images and other types of patterns in data. Geoffrey Hinton invented a method that can autonomously find properties in data, and so perform tasks such as identifying specific elements in pictures.

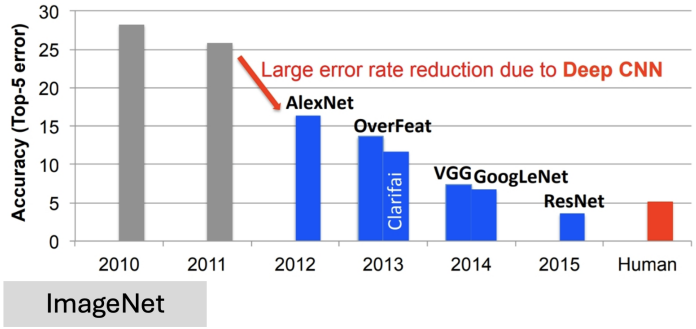
ARTIFICIAL INTELLIGENCE

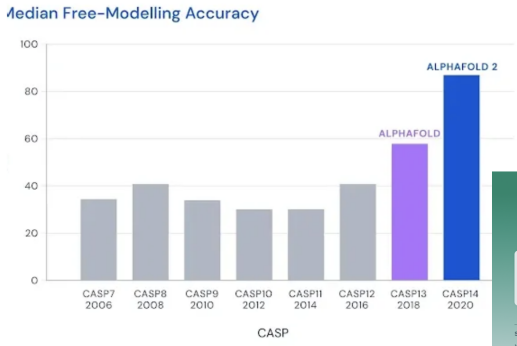
Google DeepMind leaders share Nobel Prize in chemistry for protein prediction AI

Half the prize goes to Demis Hassabis and John M. Jumper from Google DeepMind for using AI to solve protein folding, and the other to David Baker for tools to help design new proteins.

AI facts II

Outbreaking results in the last 10 years...



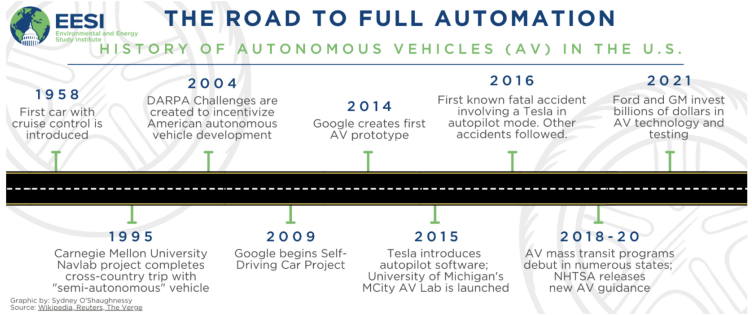


Precision Medicine

IBM Watson pinpoints rare form of leukemia after doctors misdiagnosed patient

The supercomputer identified a different type of c
the one doctors were currently treating for a patie
Japan.

By **Bernie Monegain** | August 08, 2016 | 10:58 AM



AI facts III

- AI now reached a sufficient *level of maturity* to be pervasively adopted
- Research in AI grows at an unprecedented pace and needs *significant computing and storage resources*
- The *economic impact* of AI has grown considerably (15B in 2030 EU2019)
- AI offers enormous *opportunities and challenges*, but high potential *risks and threats*

AI facts IV

- *Industry* in and with AI will change the labor market
- *Government/public administrations* make use (more and more) of AI
- AI does not only pertain scientific and technological aspects but also *social, economic, ethical and legal issues and human rights*
- *Europe* has a central and strategic role in Artificial Intelligence
- All member states have set a *National Strategy on AI* defining actions and priorities

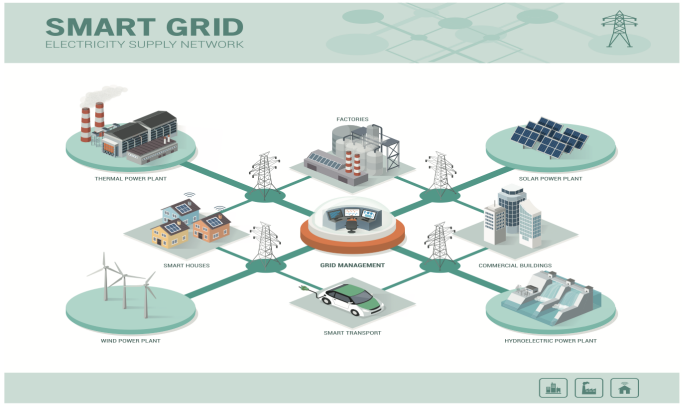
Role of AI

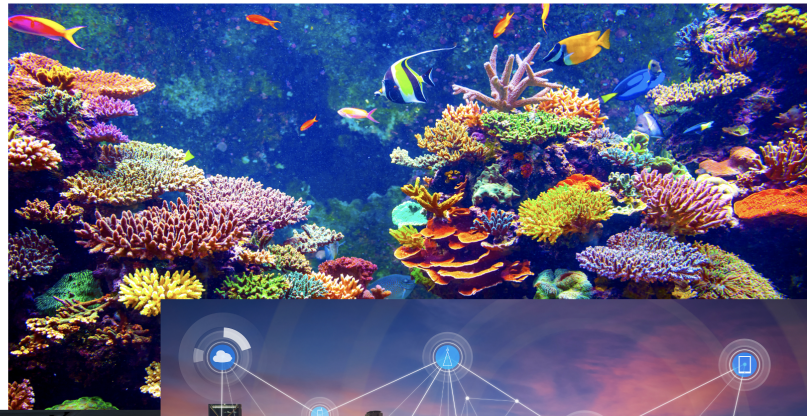
Create Models

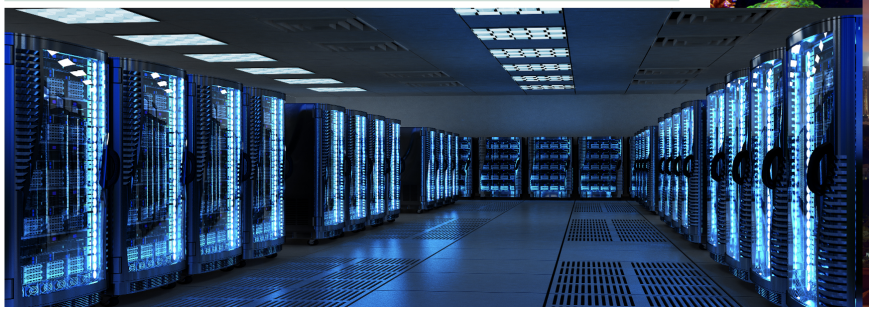
- descriptive
- predictive
- decisional
- generative

- We have seen that *AI models extracted from data* can help improve maintenance, production, energy management, retail strategies
- Do not forget that the *knowledge from experts* can be integrated into the data-driven models
- Why it is useful:
  - First we don't have to learn something we already know
  - Simplify the learning process

Impacts on Sustainability





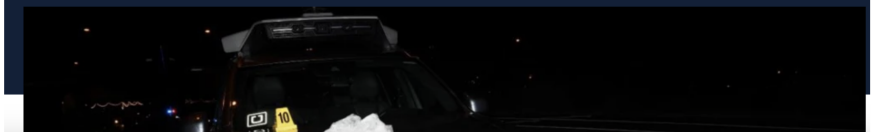


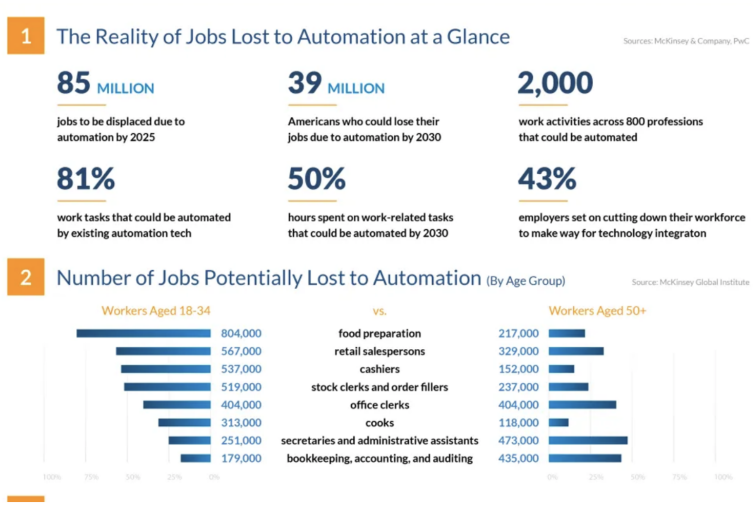
- Energy Consumption of AI Models
- Increased E-Waste
- Bias in Environmental Decision-Making
- Dependence on Non-renewable Resources

Impacts on Society

Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk

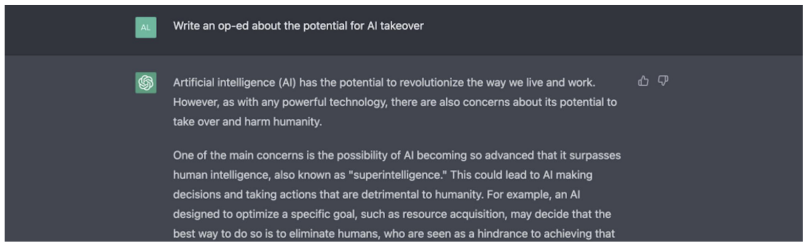
The automated car lacked "the capability to classify an object as a pedestrian unless that object was near a crosswalk," an NTSB report said.





OPINION | VIEWPOINT

ChatGPT: Exciting or terrifying?



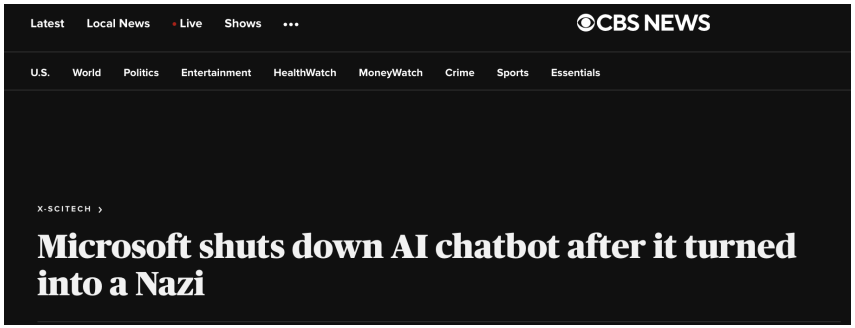
However, there are risks

- The non-responsible and not-informed use of AI can indeed bring some risks
- Risks are related to
  - the misuse of AI
  - the consequences of the use of AI without proper awareness
- AI can indeed be beneficial with applications related to health, climate change, energy, smart cities, food, equity, inclusion and sustainability at large, but it can also be applied to dangerous applications like the one on autonomous weapons, social scoring, fake news influencing public opinion.

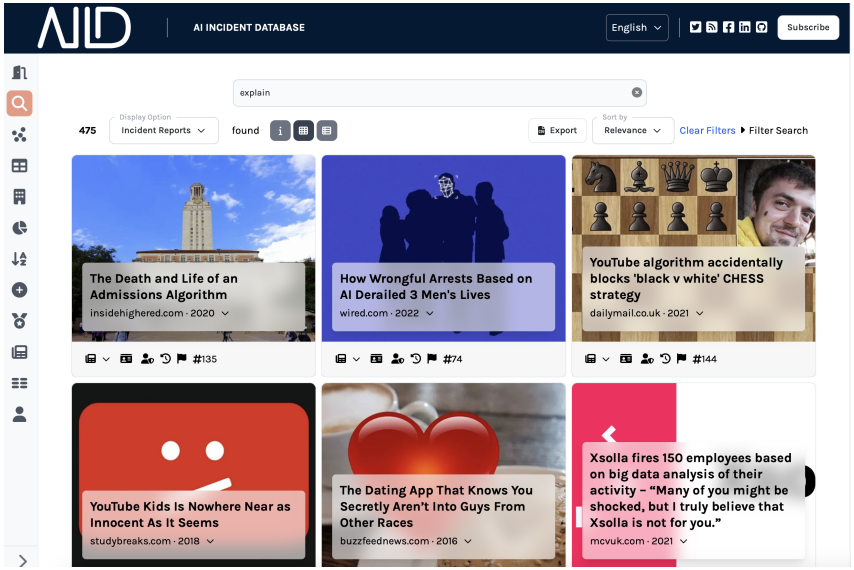
Risks: the reality I

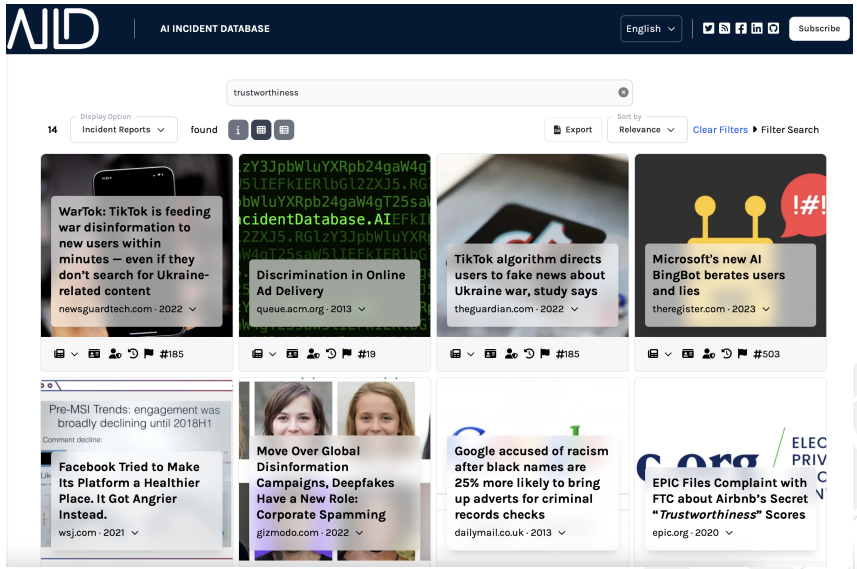
Facebook and Cambridge Analytica Election Influencing

Perhaps the most infamous example of data misuse, in 2018, news outlets revealed that the UK political consulting firm acquired and used personal data from Facebook users that was initially collected from a third party for academic research. In total, Cambridge Analytica misused the data of nearly 87 million Facebook users—many of whom had not given any explicit permission for the company to use or even access their information. Within two months of the scandal, Cambridge Analytica was bankrupt and defunct, while Facebook was left with a $5 billion fine by the Federal Trade Commission.

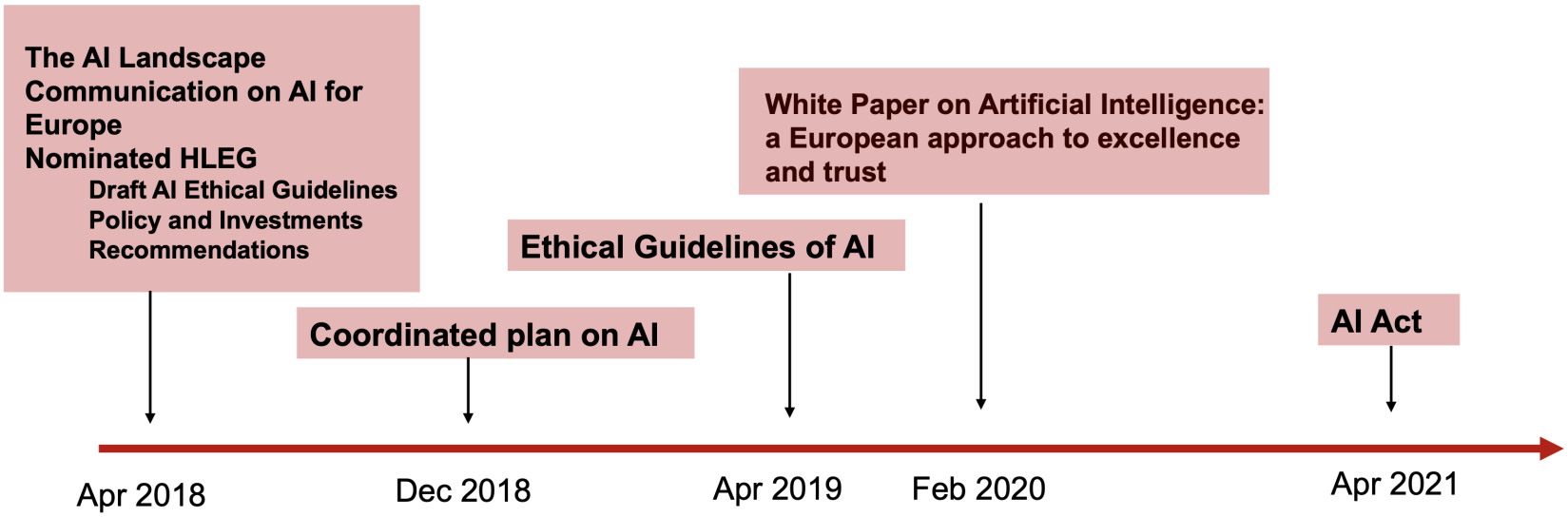


Risks: the reality II





European journey on AI



Artificial Intelligence in Europe

Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

Brussels, 25.4.2018
«Artificial Intelligence for Europe»
European Commission

…the strategy places people at the centre of the development of AI — human-centric AI. It is an approach to boost the EU's technological and industrial capacity and AI uptake across the economy, prepare for socio-economic changes, and ensure an appropriate ethical and legal framework.

Brussels, 8.4.2019
«Building Trust in Human-Centric AI»
European Commission

mobilise resources to achieve an **'ecosystem of excellence'** along the entire value chain, starting in research and innovation, and to create the right incentives to accelerate the adoption of solutions based on AI, including by small and medium-sized enterprises (SMEs).

Next in Line...

1 Ethics of AI: why?

2 Ethics of AI: initiatives

3 Ethics of AI: EGTAI

4 EGTAI Definitions & Examples

5 AI for Social Good

6 Summing up

The story so far

- AI as a field of academic dates to the 1950s [McCarthy et al., 2006]
- ethical debate is almost as old [Wiener, 1960, Samuel, 1960]
- only in recent years that impressive advances in the capabilities and applications of AI systems have brought the opportunities and risks of AI for society into sharper focus [Yang et al., 2018]

Glut of initiatives

Each additional initiative provides a supplementary statement of principles, values, or tenets to guide the development and adoption of AI (sheer volume of proposed principles—already more than 160 in 2020) [Watch, 2020])

Impression that a *me too* escalation had taken place at some point followed up by *mine and only mine*

Main Initiatives

- The Asilomar AI Principles [Morandín-Ahuerma, 2023] developed under the auspices of the Future of Life Institute in collaboration with high-level Asilomar conference in *January 2017* (hereafter Asilomar);
- The Montreal Declaration [Declaration, 2017] developed by the University of Montreal following the Forum on the Socially Responsible Development of AI in *November 2017* (hereafter Montreal);
- The general principles offered in the second version of Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems [Shahriari and Shahriari, 2017] *December 2017* (hereafter IEEE);
  - contributions from 250 global thought leaders to develop principles
  - recommendations for the ethical development and design of autonomous and intelligent systems
- The ethical principles offered in the Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems [on Ethics in Science et al., 2018] *March 2018* (hereafter EGE);
- The *five overarching principles for an AI code* [UK House of Lords report, 2018] published in *April 2018* (hereafter AIUK);
- The Tenets of the Partnership on AI [Tenets, 2018] published by a multi-stakeholder organization consisting of academics, researchers, civil society organizations, companies building and utilizing AI technology, and other groups (hereafter Partnership).

Taken together, they yield *forty-seven principles*

Main Initiatives: Comparison I

Comparison

- Differences mainly linguistic
- Degree of coherence and overlap across the six sets of principles
- Impressive and reassuring

$\Rightarrow$ convergence can be shown by comparing with the **four core principles** commonly used in **bioethics** [Beauchamp, 2016]

  - beneficence,
  - nonmaleficence,
  - autonomy, and
  - justice

Main Initiatives: Comparison II

Additional principle is needed: Explicability

- epistemological sense of intelligibility—as an answer to the question 'how does it work?'
- ethical sense of accountability—as an answer to the question 'who is responsible for the way it works?'

	Beneficence	Nonmaleficence	Autonomy	Justice	Explicability
AIUK[UK House of Lords report, 2018]	x	x	x	x	x
Asilomar[Morandín-Ahuerma, 2023]	x	x	x	x	x
EGE[on Ethics in Science et al., 2018]	x	x	x	x	x
IEEE[Shahriari and Shahriari, 2017]	x	x			x
Montreal[Declaration, 2017]	x	x	x	x	x
Partnership[Tenets, 2018]	x	x		x	x
AI4People	x	x	x	x	x
HLEG	x	x	x	x	x
OECD	x	x	x	x	x
Beijing	x	x		x	x
Rome Call	x	x	x	x	x

Next in Line...

1 Ethics of AI: why?

2 Ethics of AI: initiatives

3 Ethics of AI: EGTAI

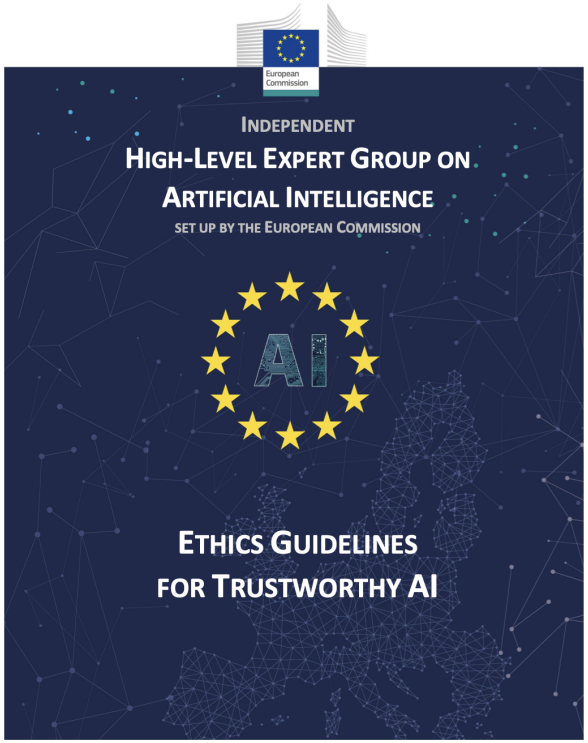
4 EGTAI Definitions & Examples

5 AI for Social Good

6 Summing up

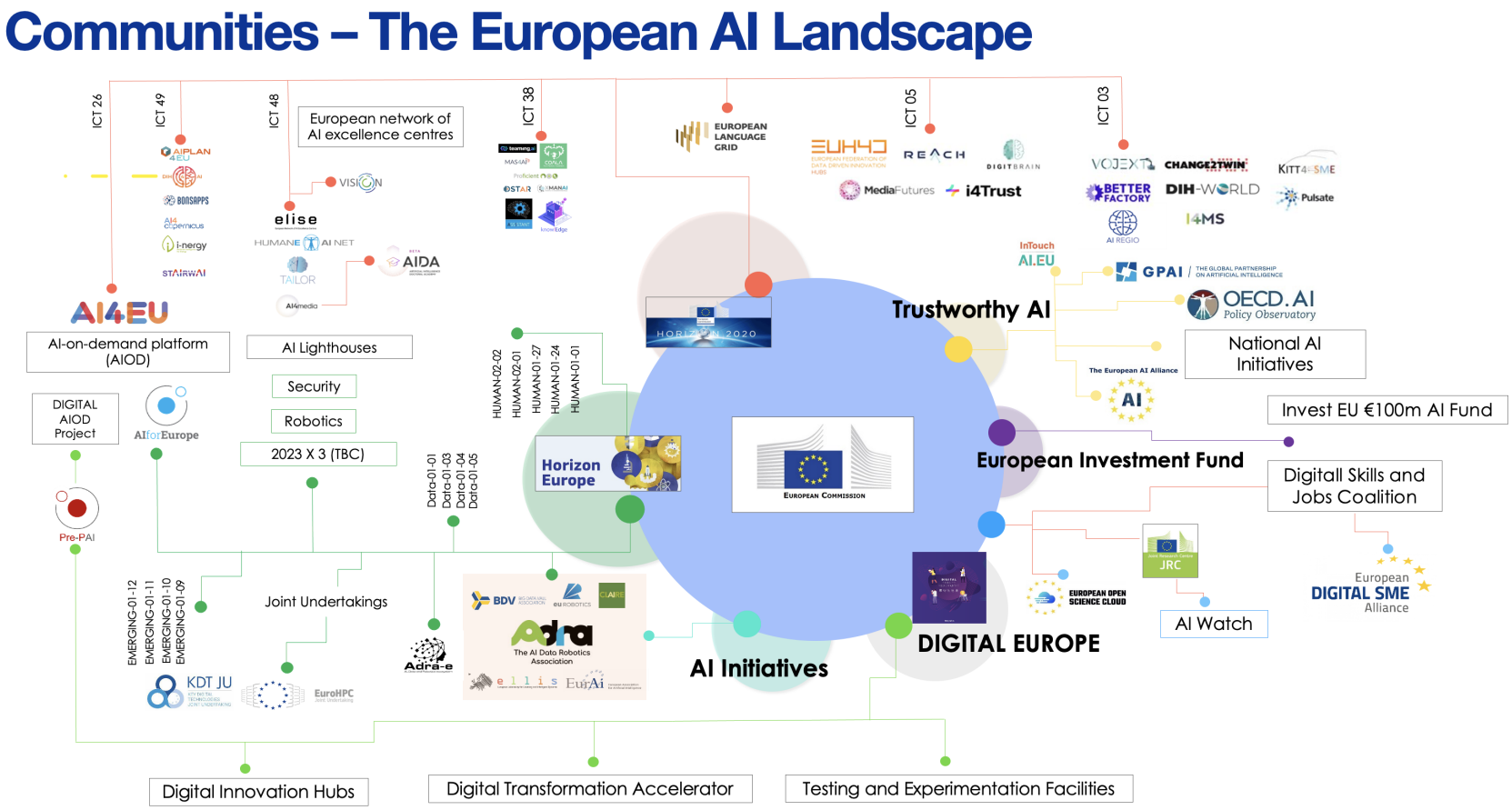
EGTAI

Ethical guidelines of Trustworthy AI



- Important document defining dimensions of AI for ensuring trust
- The base for the white paper and the AI act

EGTAI: Motivation I



EGTAI: Motivation II

AI ACT

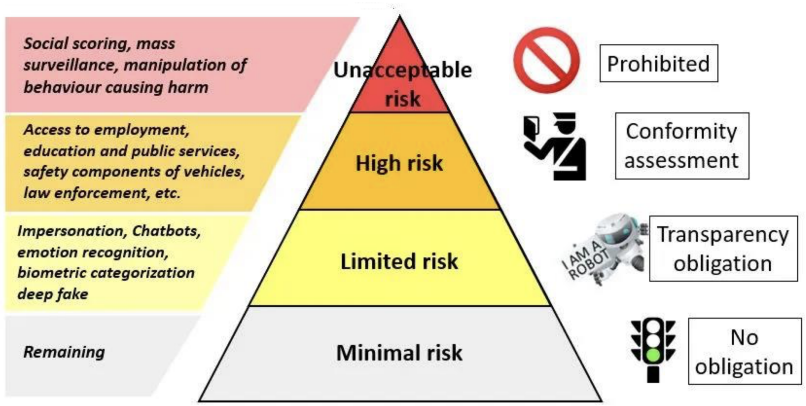
If AI produces harm, understand who is responsible for it

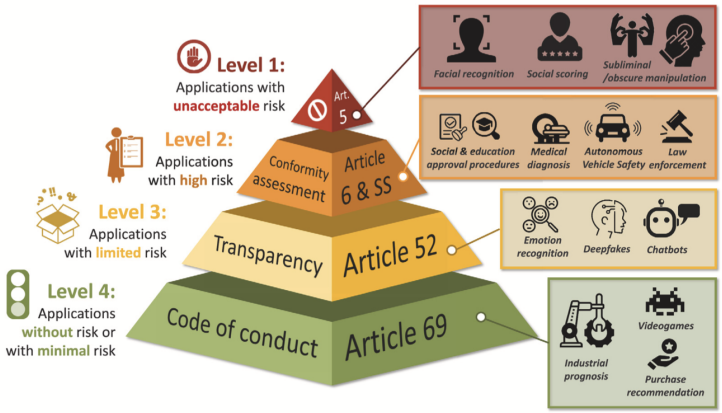
- Writer of the algorithm
- Owner of the object in which AI algorithm works (the car for example)
- The human supervising it

Legal aspects involved.

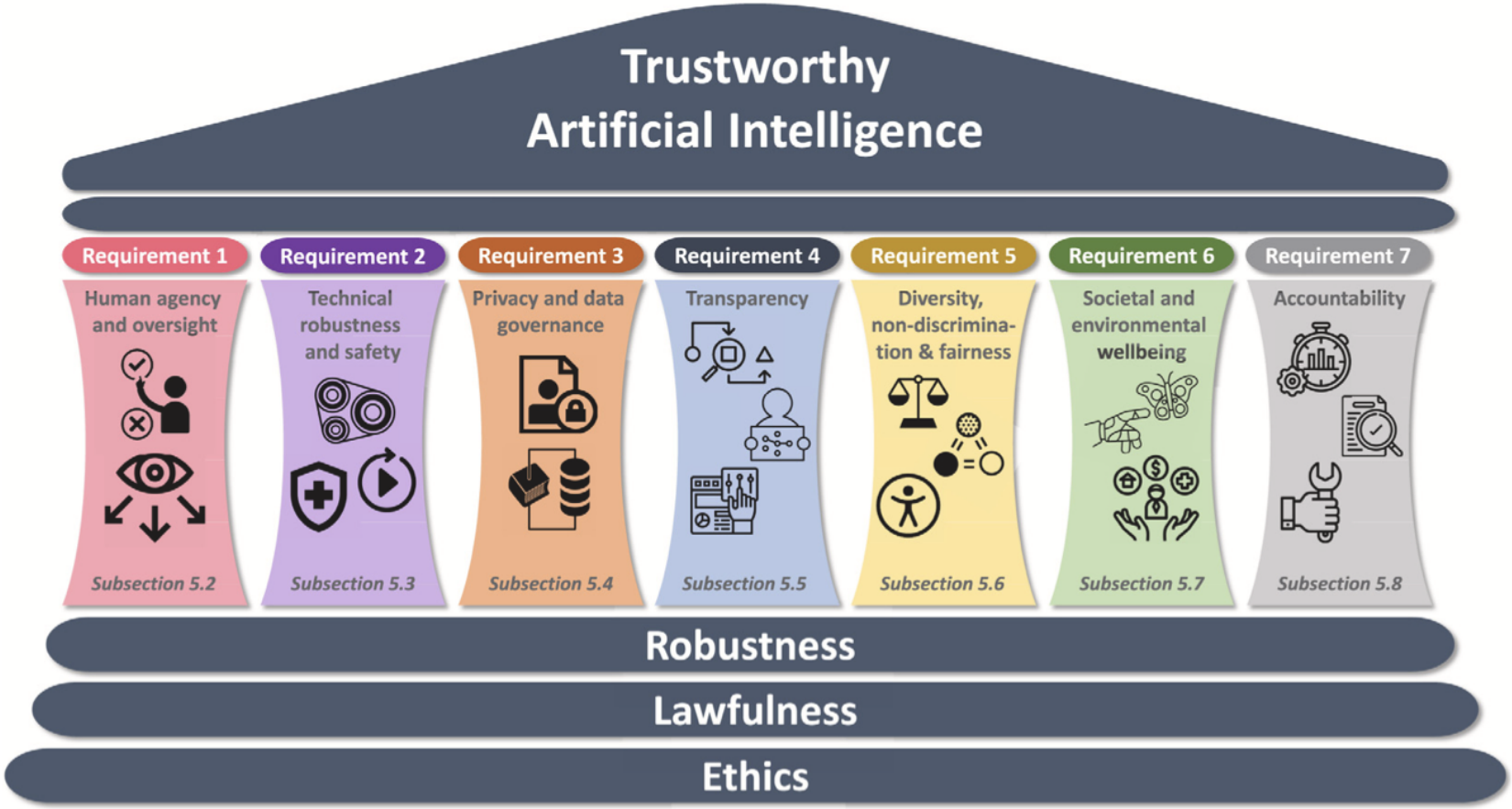
EGTAI: Motivation III

AI ACT: AI RISK LEVELS

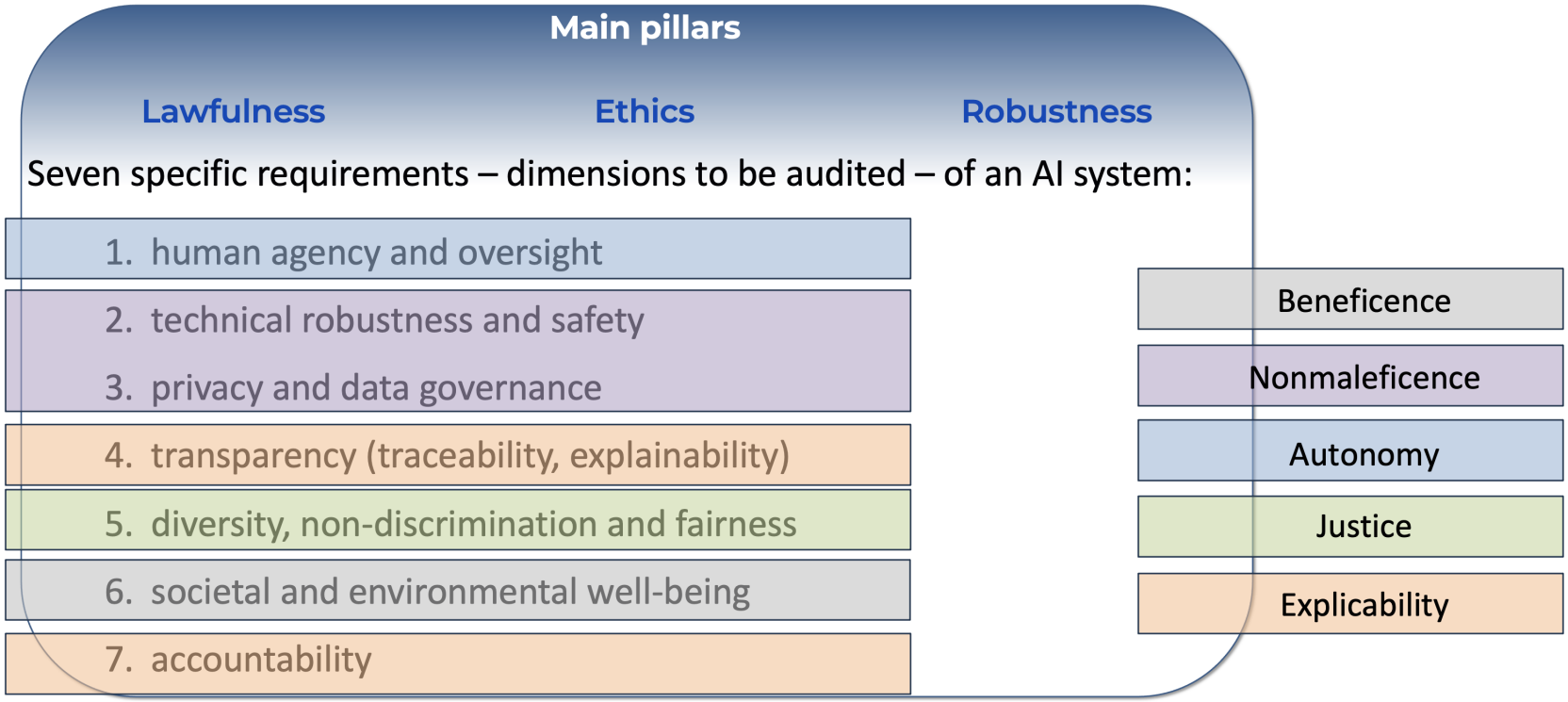




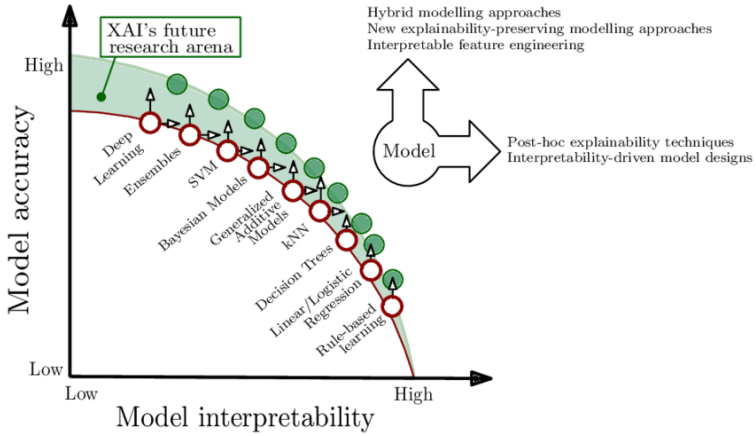
EGTAI: in Detail I

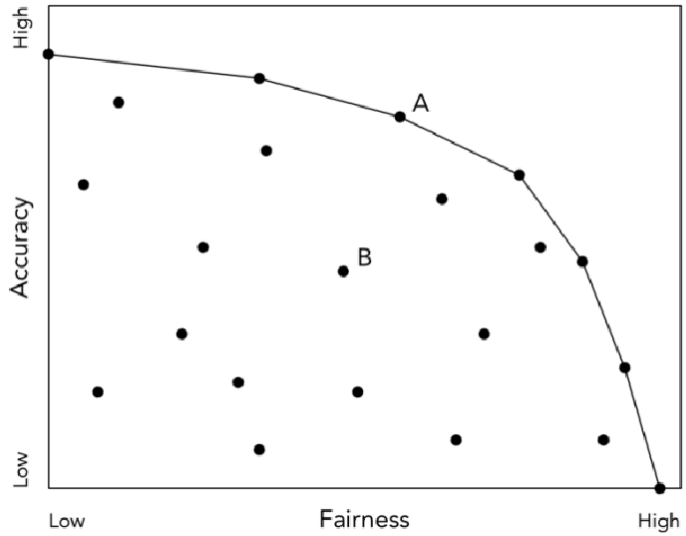


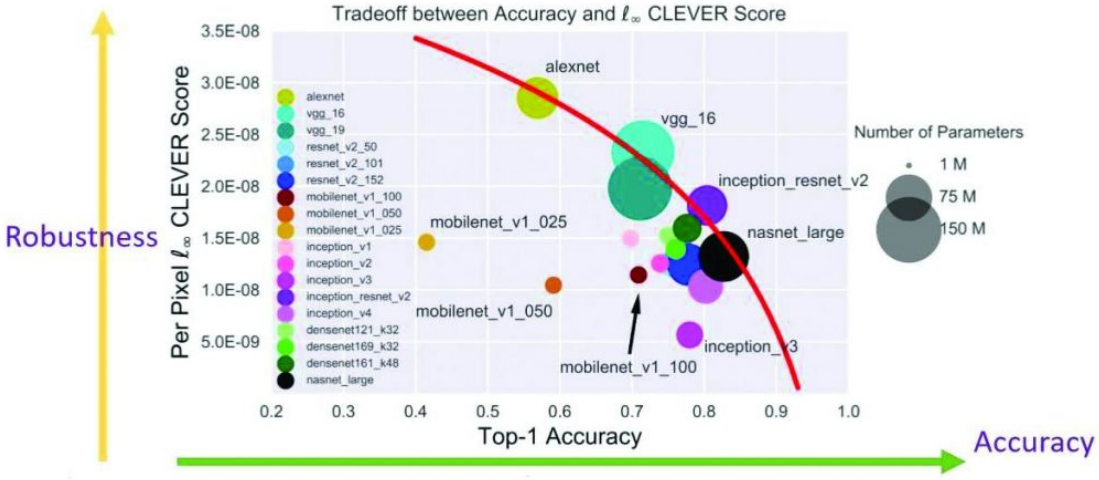
EGTAI: in Detail II



EGTAI & Trade Offs







Next in Line...

1 Ethics of AI: why?

2 Ethics of AI: initiatives

3 Ethics of AI: EGTAI

4 EGTAI Definitions & Examples

5 AI for Social Good

6 Summing up

Human agency and oversight I

Human agency

Empower human beings, allowing them to make informed decisions and fostering their fundamental rights. Agency may be achieved through governance mechanisms such as a *human-centric approaches*, *AI for social good*, *human computation*, *interactive machine learning*

Human oversight

Ensuring that an AI system does not undermine human autonomy or cause other adverse effects. Oversight may be achieved through governance mechanisms such as a *human-in-the-loop*, *human-on-the-loop*, *human-in-command*

Human agency and oversight II

The European Commission is promoting human-centered technology

- The technology is built and designed around humans
- Human needs are taken into account in the design process
- Smooth interaction between humans and technology

Technical robustness and safety I

AI Safety

Prevention of unintentional harm (physical or psychological) especially to human beings, and by extension to other material or immaterial elements that may be valuable for humans, including the system itself and the *minimization of the consequences of intentional harm*. These include

- resilience of AI-based systems (to attacks and security)
- ensuring fallback plans (in case something goes wrong)
- general safety, and being *accurate*, *reliable* and *reproducible*
- cover the way and conditions in which the system ceases its operation, and the consequences of stopping

Technical robustness and safety II

Robustness

Emphasises that safety and —conditionally to it— *functionality, must be preserved in different situations and also under harsh conditions*, including unanticipated errors, exceptional situations, unintended or intended damage, manipulation or catastrophic states.

Technical robustness and safety III

Reproducibility

Once robustness and safety have been addressed, an important dimension in this key requirement for trustworthy AI is reproducibility.

- *Repeatability* (same team, same experimental setup), which means that an individual or a team of individuals can reliably repeat his/her/their own experiment.
- *Replicability* (different team, same experimental setup): an independent group of individuals can obtain the same result using artifacts that they independently develop in their entirety.
- *Reproducibility* (different team, different experimental setup with stated precision): a different independent group can obtain the same result using their own artifacts.

Technical robustness and safety IV

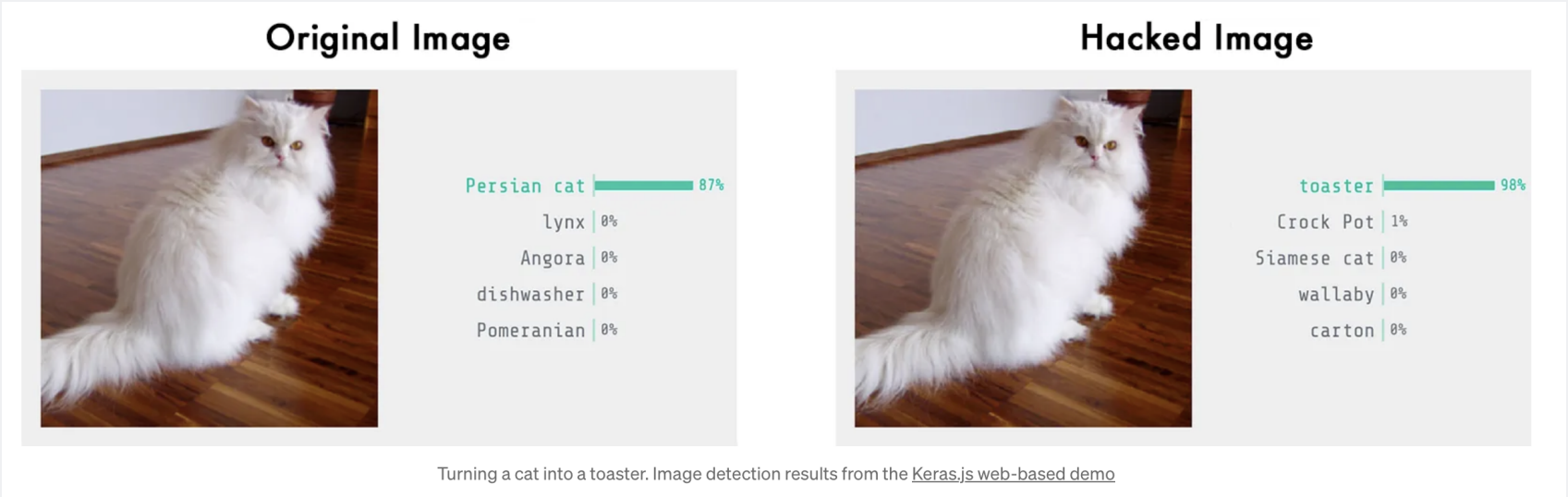
Safety concerns driving without accidents neither for the driver nor for other vehicles, pedestrians, bicycles. . . .

...Self-driving cars should be safe even under conditions as fog, rain, snow, heavy traffic

Technical robustness and safety V

NN can be hacked

DNN can be unstable: by applying some perturbations to inputs we can arbitrarily change the output



Turning a cat into a toaster. Image detection results from the Keras.js web-based demo

Example from https://medium.com/@ageitgey/machine-learning-is-fun-part-8-how-to-intentionally-trick-neural-networks-

Privacy and data governance I

Privacy and data governance

This is the most well-known and better-regulated aspect of AI

- GDPR has been around for quite some years
- European Union strategy for data proposed the Data Act (2022) a regulation harmonizing rules on fair access to and use of data. In practice this regulation complements the Data Governance Act by specifying who can create value from data and under which circumstances

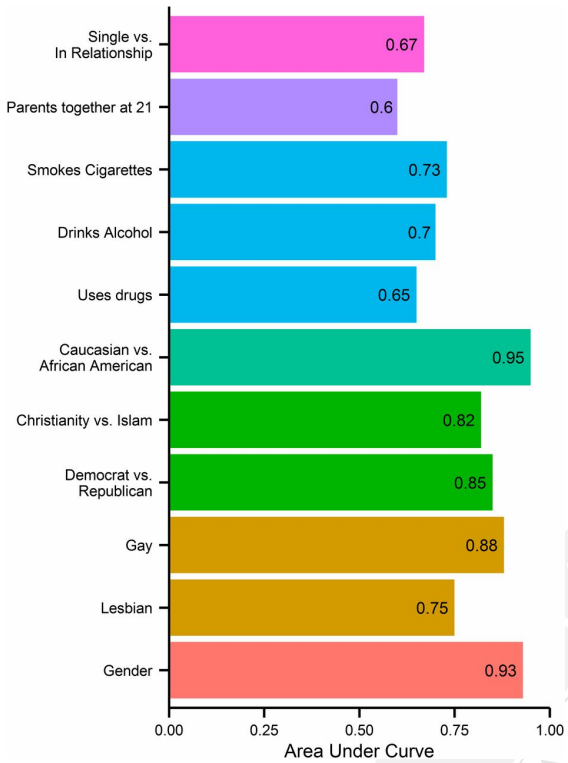
Here there are a couple of aspects that might not be so well known:

- Data might contain information we do not expect
- *IPR might become an issue*

Privacy and data governance II

Easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender

- A dataset of 58,000 volunteers has made available their Facebook Likes and detailed personal data, profiles etc. for machine learning
- The learned model is accurate and discriminates among different categories (homosexual and heterosexuality with 88% accuracy, Democratics and Repubblicans with 85% accuracy)
- Implications on privacy (Cambridge Analytica)



https://www.pnas.org/doi/full/10.1073/pnas.1218772110

Privacy and data governance III

- Generative AI is trained with real data and generates similar synthetic data.
- *Who owns the generated output?*
- This holds for new music, new paintings, poetry, choreographies, movies

Award Winning Picture by AI, Boris Eldagsen

Transparency I

Transparency

Ensure appropriate information reaches the relevant stakeholders.

Humans must be informed of systems' capabilities and limitations and always be aware that they are interacting with AI systems.

- *explanations* should be timely, adapted and communicated to the stakeholder audience concerned
- *traceability* of AI systems should be ensured
- *communication*

Transparency II

Not common definition of explanation: this is the most widely accepted (NIST report)

Explanation

An explanation is the evidence, support, or reasoning related to a system's output or process, where the output of a system differs by task, and the process refers to the procedures, design, and system workflow which underlie the system.

Transparency III

Explanation — Communication

The explanations returned depend on various factors, such as:

- the type of task they are needed for,
- on which kind of data the AI system acts,
- who is the final user of the explanation,
- if they allow to explain the whole behavior of the AI system (global explanation) or reveal the reasons for the decision only for a particular instance (local explanation)
- the business perspective, i.e., which are the implications of companies in having explainable and interpretable systems and models, in terms of business strategies and secrecy,
- the fact that, in a decentralized node, an explanation could require information that is not directly available on site.

Transparency IV

- Bias in – bias out: system trained for distinguishing photos of Wolves and Eskimo Dogs (huskies).
- Training phase 20 images, hand-selected such that all pictures of wolves had snow in the background, while pictures of huskies did not.
- The algorithm predicts Wolves every time there is snow, Eskimo Dogs otherwise with this explanation

(a) Husky classified as wolf

(b) Explanation

Diversity, non-discrimination and fairness I

Diversity, non-discrimination, accessibility, universal design and stakeholder participation

- inclusion of diverse data and people, and ensures that individuals at risk of exclusion have equal access to AI benefits
- diversity, it advocates for the need for heterogeneous and randomly sampling procedures for data acquisition, diverse representation of a population that includes minorities, and the assurance for non-discriminating automated processes that lead to unfairness or biased models

Diversity, non-discrimination and fairness II

Fairness — Article 21 of the EU Charter of Fundamental Rights

> *any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.*

They describe two different discrimination scenarios:

1. direct discrimination (disparate *treatment*)
2. indirect discrimination (disparate *impact*): when a seemingly "neutral provision, criterion or practice" disproportionately disadvantages members of a given sensitive group compared to others

Diversity, non-discrimination and fairness III

Computational fairness

- potential biases and discrimination that can arise from the use of computational algorithms
- ensuring algorithms do not perpetuate or amplify existing biases and do not discriminate against certain groups of people based on sensitive attributes

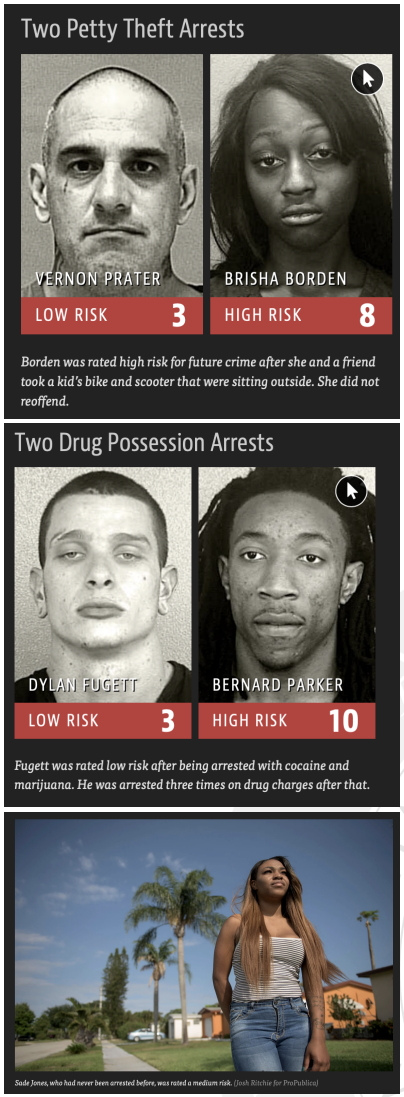
Fairness Metrics

Quantitative measurement used to assess and quantify the fairness or bias of an algorithm's predictions or decisions

Diversity, non-discrimination and fairness IV

Compas

- COMPAS tries to predict, among other indexes, the recidivism of defendants, who are ranked low, medium, or high risk
- Used in many US states (such as New York and Wisconsin)
- Accuracy around 70
- ProPublica journalists found that *black defendants* were far *more likely* than *white defendants* to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk



Diversity, non-discrimination and fairness V

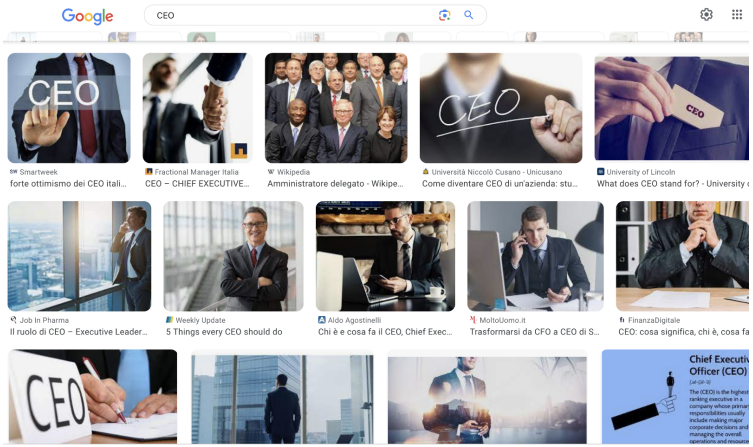
Amazon Recruiting System

- shut down its experimental AI recruiting tool after discovering it *discriminated against women* $\rightarrow$ tool created to trawl the web and spot potential candidates, rating them from one to five stars
- algorithm learned to systematically downgrade women's CV's for technical jobs such as software developer

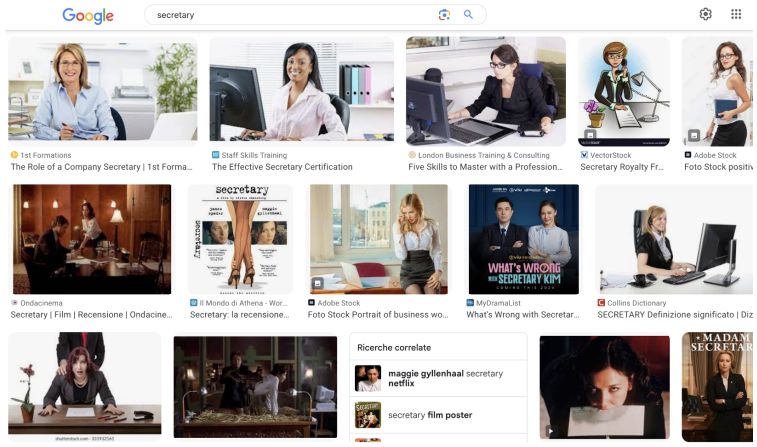
$\Rightarrow$ Reason: data used to train the system were data from the last 10 years of recruitment $\rightarrow$ manly man profiles

Diversity, non-discrimination and fairness VI

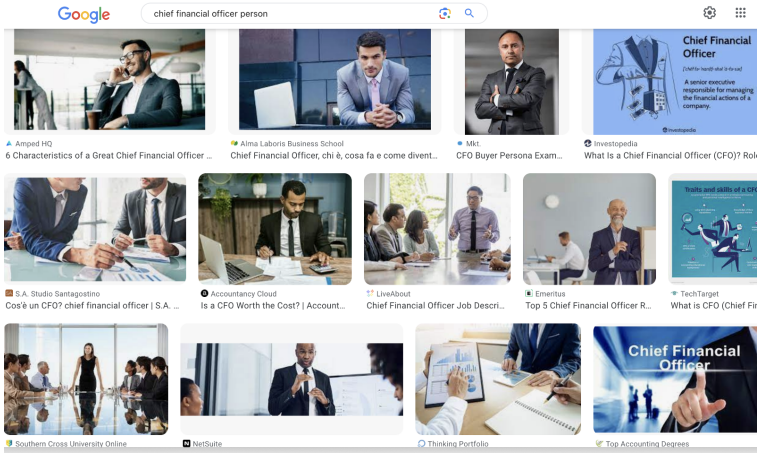
CEO Images from the web



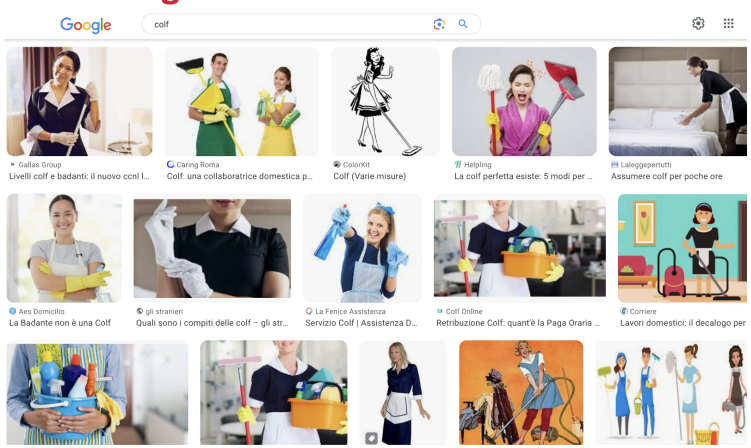
SECRETARY Images from the web



CFO Images from the web



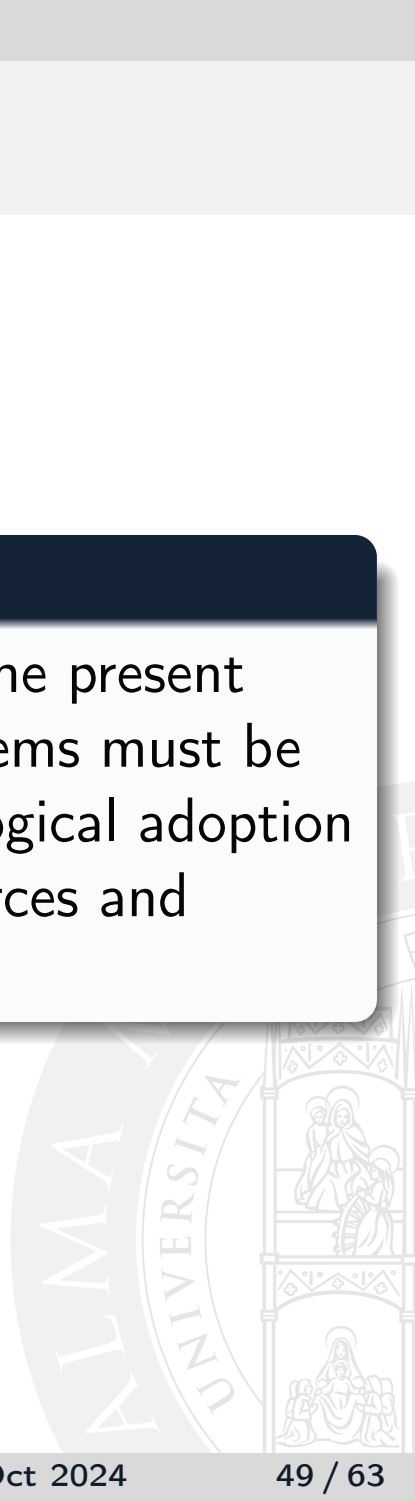
COLF Images from the web



Societal and environmental wellbeing I

Societal and environmental wellbeing

AI-based systems should benefit all humankind, not only at the present time but also in future generations. Therefore, AI-based systems must be *sustainable and environmentally friendly*, so that the technological adoption of AI does not entail a progressive depletion of natural resources and maintains an ecological balance



Societal and environmental wellbeing II

Sustainability and environmental wellbeing

- AI needs High performance computing that has a heavy carbon footprint
- AI can help reducing this carbon footprint
  - Power aware management
  - Optimal allocation and scheduling
  - Cooling optimization and thermal-aware workload dispatching
  - Anomaly detection

Societal and environmental wellbeing III

Societal wellbeing

- AI can improve social welfare
  - perform routine tasks in an autonomous safer, and more efficient fashion, enhancing productivity and improving the quality of life of humankind
  - speed up processes, smooth administrative bottlenecks and save paperwork
  - help city planners, e.g., by visualizing the consequences of climate change, predicting future floods
  - ...

Societal and environmental wellbeing IV

High-Performance Computing (HPC) power consumption

- Today supercomputers are power limited $\rightarrow$ need for improved energy efficiency. . . ..
  - Acceptable range for an Exascale supercomputer is 20MWatts
  - *Power Capping* is a widespread method to deal with the increasing energy-related concerns
  - Idea: constrain a supercomputer power consumption within a certain power budget

Accountability I

Accountability

ensure responsibility and accountability for the development, deployment, maintenance and/or use of AI systems and their outcomes

Accountability II

If AI produces harm, understand who is responsible for it

- Writer of the algorithm
- Owner of the object in which AI algorithm works (the car for example)
- The human supervising it

Legal aspects involved

Accountability III

Auditability

- development of practical tools capable of verifying desirable properties of neural networks such as *stability*, *sensitivity*, *relevance* or *reachability*
- as well as metrics beyond explainability, such as *traceability*, *data quality* and *integrity*

Auditability is becoming increasingly important when standards are being materialized touching upon all AI requirements: IEEE, ISO/IEC and CEN/CENELEC, which are implementing concrete guidelines to apply trustworthy AI requirements in industrial setups

Next in Line...

1 Ethics of AI: why?

2 Ethics of AI: initiatives

3 Ethics of AI: EGTAI

4 EGTAI Definitions & Examples

5 AI for Social Good

6 Summing up

AI for Social Good

Helps to reduce, mitigate, or eradicate a given social or environmental problem without introducing new harms or amplifying existing ones.

Next in Line...

1. Ethics of AI: why?
2. Ethics of AI: initiatives
3. Ethics of AI: EGTAI
4. EGTAI Definitions & Examples
5. AI for Social Good
6. Summing up

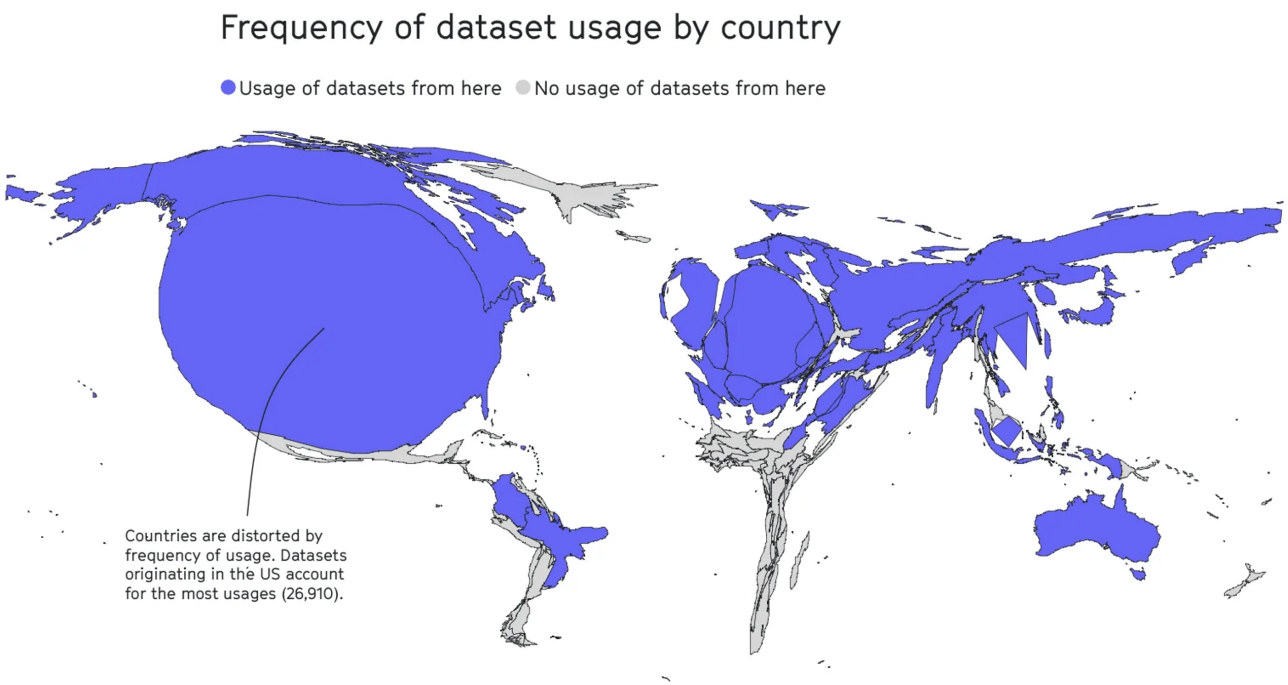
Next...

- Fairness in AI — A computational perspective
- Explainability — A computational perspective

AI: More is Better? I

- More data,
- More algorithms,
- More computational power.

Data: How AI sees the world

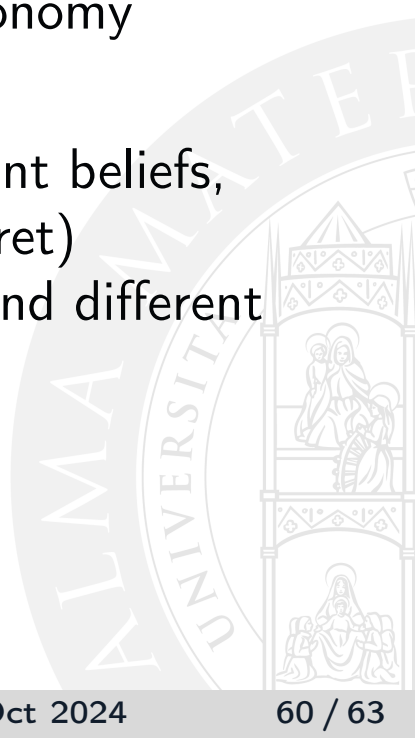


This map shows how often 1,933 datasets were used (43,140 times) for performance benchmarking across 26,535 different research papers from 2015 to 2020.

Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research, Bernard Koch, Emily Denton, Alex Hanna, Jacob G. Foster, 2021. Map made with Natural Earth. Distorted with cartogram3.

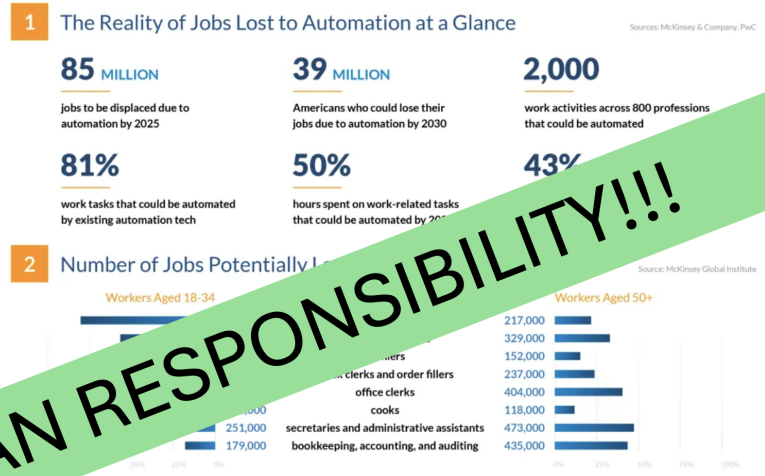
Algorithms: How AI models the world?

- AI is a *rational* system
  - AI agents have *preferences, or priorities, on outcomes of actions*;
  - AI agents *optimize* actions based on those preferences
- AI principles are *Global North stereotypes*
  - Optimisation / Efficiency / Rationality / Agency / Autonomy
- But... *People act in context*
  - pursue seemingly incompatible goals and hold inconsistent beliefs, (e.g., altruism, fairness, justice, or to prevent future regret)
  - actions are influenced by the context, including others and different situations.



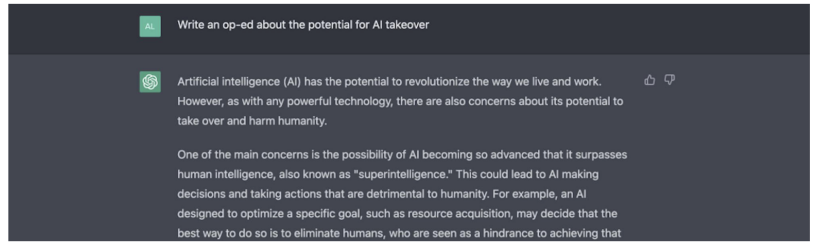
Self-driving Uber car that hit woman did not recognise pedestrians jaywalk

The automat… classify an object as a pedestrian unless that object w… it said.



OPINION | VIEWPOINT

ChatGPT: Exciting or terrifying?



IMPACT MEANS HUMAN RESPONSIBILITY!!!

Foundations of Ethical Artificial Intelligence

Foundations of Ethical Artificial Intelligence

Computational Perspective

Roberta Calegari
roberta.calegari@unibo.it

Alma Mater Studiorum – Università di Bologna

10 October 2024

References I

[Beauchamp, 2016] Beauchamp, T. L. (2016).
Principlism in bioethics.
Bioethical decision making and argumentation, pages 1–16
.

[Declaration, 2017] Declaration, M. (2017).
For a responsible development of artificial intelligence
.

[McCarthy et al., 2006] McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (2006).
A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955.
AI magazine, 27(4):12–12
.

[Morandín-Ahuerma, 2023] Morandín-Ahuerma, F. (2023).
Twenty-three asilomar principles for artificial intelligence and the future of life
.

[on Ethics in Science et al., 2018] on Ethics in Science, E. G., Technologies, N., et al. (2018).
Statement on artificial intelligence, robotics and'autonomous' systems: Brussels, 9 March 2018.
EU: European Union
.

[Samuel, 1960] Samuel, A. L. (1960).
Some moral and technical consequences of automation—a refutation.
Science, 132(3429):741–742
.

References II

[Shahriari and Shahriari, 2017] Shahriari, K. and Shahriari, M. (2017).
Ieee standard review—ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems.
In *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, pages 197–201. IEEE
.

[Tenets, 2018] Tenets (2018).
Partnership on ai
.

[UK House of Lords report, 2018] UK House of Lords report (2018).
Ai in the uk: ready, willing and able? (aiuk).
https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf
.

[Watch, 2020] Watch, A. (2020).
Ai ethics guidelines global inventory.
https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/
.

[Wiener, 1960] Wiener, N. (1960).
Some moral and technical consequences of automation: As machines learn they may develop unforeseen strategies at rates that baffle their programmers.
Science, 131(3410):1355–1358
.

References III

[Yang et al., 2018] Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., et al. (2018).
The grand challenges of science robotics.
Science robotics, 3(14):eaar7650
.