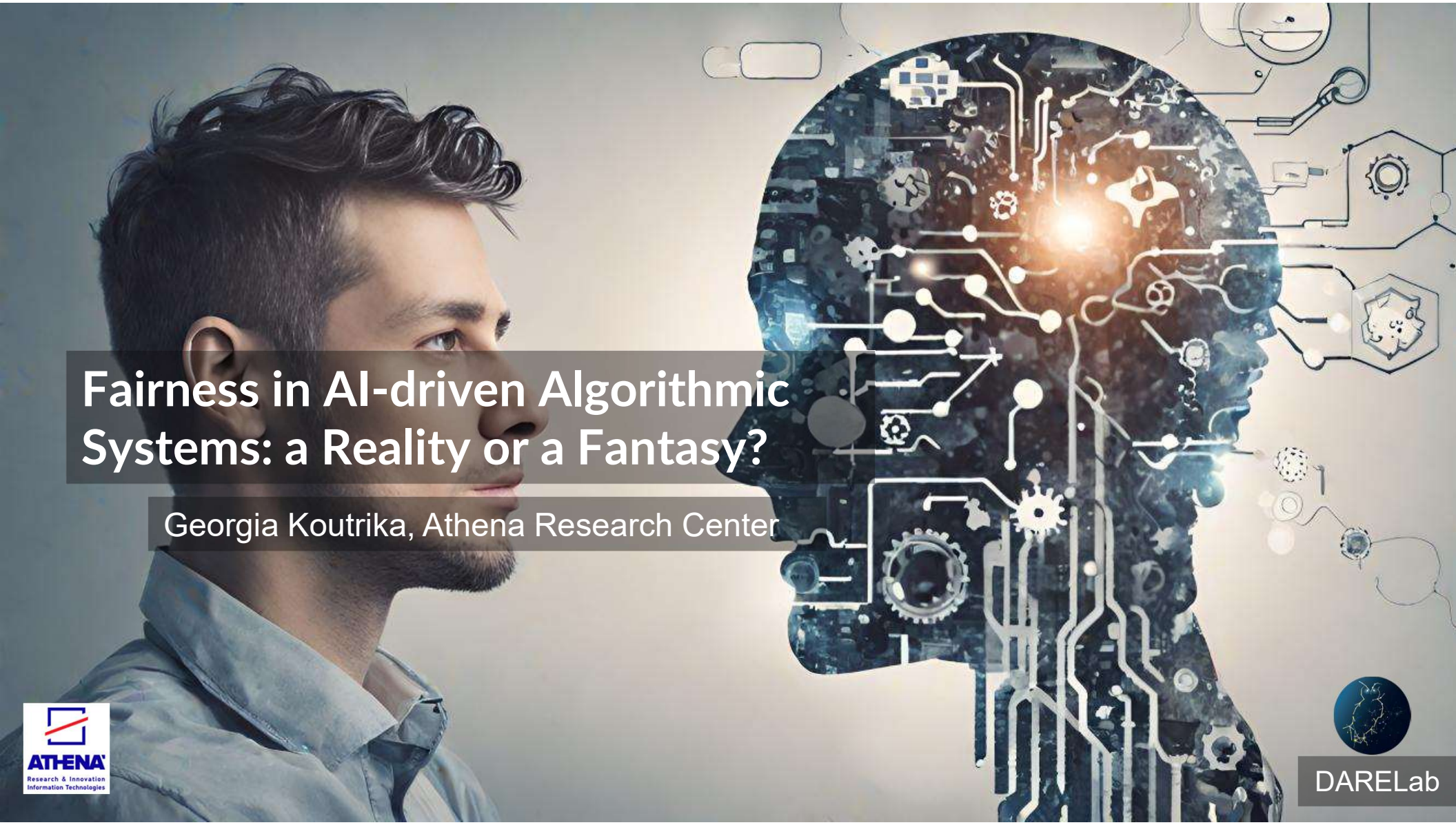


Fairness in AI-driven Algorithmic Systems: a Reality or a Fantasy?

Georgia Koutrika, Athena Research Center

ATHENA Research & Innovation Information Technologies

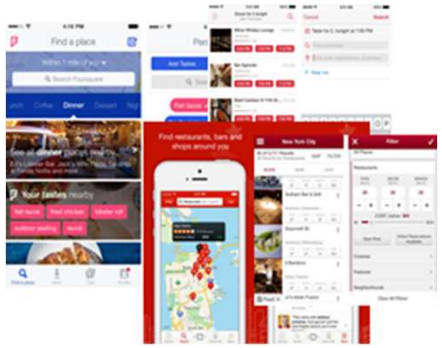
DARELab

Algorithmic fairness: why?

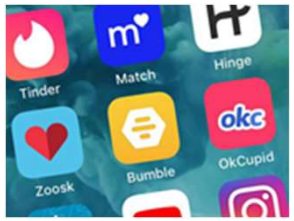
We live in a world where decisions are assisted or taken by algorithmic systems driven by large amounts of data

From simple (or not that simple) personal decisions

Where to eat?

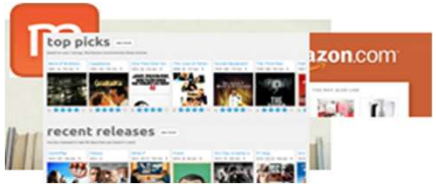


Who to date?



What are the news?

What to read, watch, buy..?



Algorithmic fairness: why?

To more complex ones ...

- Education, school admission
- Law enforcement, sentencing decisions
- Job recruitment
- Insurance, Credit
- Housing
- Pricing of goods and services
- ...

How much can/should we trust such systems?

And this concern has not been without reason:

a steady stream of empirical findings has shown that data-driven methods can unintentionally both encode **existing human biases** and **introduce new ones**.

A Snapshot of the Frontiers of Fairness in Machine Learning
By Alexandra Chouldechova, Aaron Roth
Communications of the ACM, May 2020, Vol. 63 No. 5, Pages 82-89

Job Ads

Facebook disproportionately shows certain types of job ads to men and women.
It was more likely to present job ads to users if **their gender identity reflected the concentration of that gender in a particular position** or industry

(study led by University of Southern California researchers)

Example:

Ads for delivery driver job listings that had similar qualification requirements but for different companies.

- The ads did not specify a specific demographic.
- One was an ad for Domino's pizza delivery drivers, the other for Instacart drivers.
- Instacart has more female drivers but Domino's has more male drivers.

Facebook targeted the **Instacart delivery job to more women** and the **Domino's delivery job to more men**.

https://www.wsj.com/articles/facebook-shows-men-and-women-different-job-ads-study-finds-11617969600

Music streaming recommendations

Music streaming recommendations

Diversity problem:
E.g., Spotify's prime 10 most streamed artists of 2020 include just two female singers

Rank	Song	Artist(s)	Album	Streams (millions)	Date published	Ref(s)
1	"Shape of You"	Ed Sheeran	÷	2,768	6 January 2017	[11]
2	"Dance Monkey"	Tones and I	*The Kids Are Coming*	2,168	10 May 2019	[12]
3	"Rockstar"	Post Malone featuring 21 Savage	*Beerbongs & Bentleys*	2,159	15 September 2017	[13]
4	"Blinding Lights"	The Weeknd	*After Hours*	2,152	29 November 2019	[14]
5	"One Dance"	Drake featuring Wizkid and Kyla	*Views*	2,008	5 April 2016	[15]
6	"Closer"	The Chainsmokers featuring Halsey	*Collage*	1,989	29 July 2016	[16]
7	"Sunflower"	Post Malone and Swae Lee	*Spider-Man: Into the Spider-Verse* and *Hollywood's Bleeding*	1,881	18 October 2018	[17]
8	"Someone You Loved"	Lewis Capaldi	*Breach* and *Divinely Uninspired to a Hellish Extent*	1,868	8 November 2017	
9	"Señorita"	Shawn Mendes and Camila Cabello	*Shawn Mendes* and *Romance*	1,798	21 June 2019	[18] [19]
10	"Thinking Out Loud"	Ed Sheeran	×	1,742	20 June 2014	[20]
11	"Bad Guy"	Billie Eilish	*When We All Fall Asleep, Where Do We Go?*	1,739	29 March 2019	[21]
12	"God's Plan"	Drake	*Scorpion*	1,724	19 January 2018	[22]
13	"Say You Won't Let Go"	James Arthur	*Back from the Edge*	1,697	9 September 2016	
14	"Perfect"	Ed Sheeran	÷	1,689	3 March 2017	[23]
15	"Believer"	Imagine Dragons	*Evolve*	1,673	1 February 2017	
16	"Lucid Dreams"	Juice Wrld	*Goodbye & Good Riddance*	1,600	4 May 2018	
17	"Havana"	Camila Cabello featuring Young Thug	*Camila*	1,583	3 August 2017	[24]
18	"Photograph"	Ed Sheeran	×	1,581	20 June 2014	[25]
19	"Starboy"	The Weeknd featuring Daft Punk	*Starboy*	1,557	21 September 2016	[26]
20	Sad!"	XXXTentacion	?	1,556	1 March 2018	[27]
21	"Love Yourself"	Justin Bieber	*Purpose*	1,546	9 November 2015	[28]
22	"Something Just Like This"	The Chainsmokers and Coldplay	*Memories...Do Not Open*	1,522	22 February 2017	[29]
23	"New Rules"	Dua Lipa	*Dua Lipa*	1,500	2 June 2017	
24	"Thunder"	Imagine Dragons	*Evolve*	1,478	27 April 2017	
25	"Bohemian Rhapsody"	Queen	*A Night at the Opera*	1,466	31 October 1975	
26	"Lean On"	Major Lazer and DJ Snake featuring MØ	*Peace Is the Mission*	1,465	2 March 2015	
27	"Shallow"	Lady Gaga and Bradley Cooper	*A Star Is Born*	1,462	27 September 2018	
28	"XO Tour Llif3"	Lil Uzi Vert	*Luv Is Rage 2*	1,456	24 March 2017	
29	"All of Me"	John Legend	*Love in the Future*	1,447	12 August 2013	
30	"Sorry"	Justin Bieber	*Purpose*	1,439	22 October 2015	
31	"Despacito (Remix)"	Luis Fonsi and Daddy Yankee featuring Justin Bieber	*Vida*	1,436	17 April 2017	
32	"Happier"	Marshmello and Bastille	Non-album single	1,427	17 August 2018	
33	"7 Rings"	Ariana Grande	*Thank U, Next*	1,423	18 January 2019	
34	Humble"	Kendrick Lamar	*Damn*	1,414	30 March 2017	[30]
35	"Jocelyn Flores"	XXXTentacion	*17*	1,413	25 August 2017	[31]

- When a recommendation service gives suggestions, it does so by leveraging the historical data.
- That creates a suggestion loop, **amplifying existing bias** and **reducing diversity**.

https://digitpatrox.com/not-ok-computer-music-streamings-diversity-problem/

Word-2vec embedding

Extreme *she* occupations		
1. homemaker	2. nurse	3. receptionist
4. librarian	5. socialite	6. hairdresser
7. nanny	8. bookkeeper	9. stylist
10. housekeeper	11. interior designer	12. guidance counselor

Extreme *he* occupations		
1. maestro	2. skipper	3. protege
4. philosopher	5. captain	6. architect
7. financier	8. warrior	9. broadcaster
10. magician	11. figher pilot	12. boss

The most extreme occupations as projected on to the *she*–*he* gender direction on g2vNEWS.

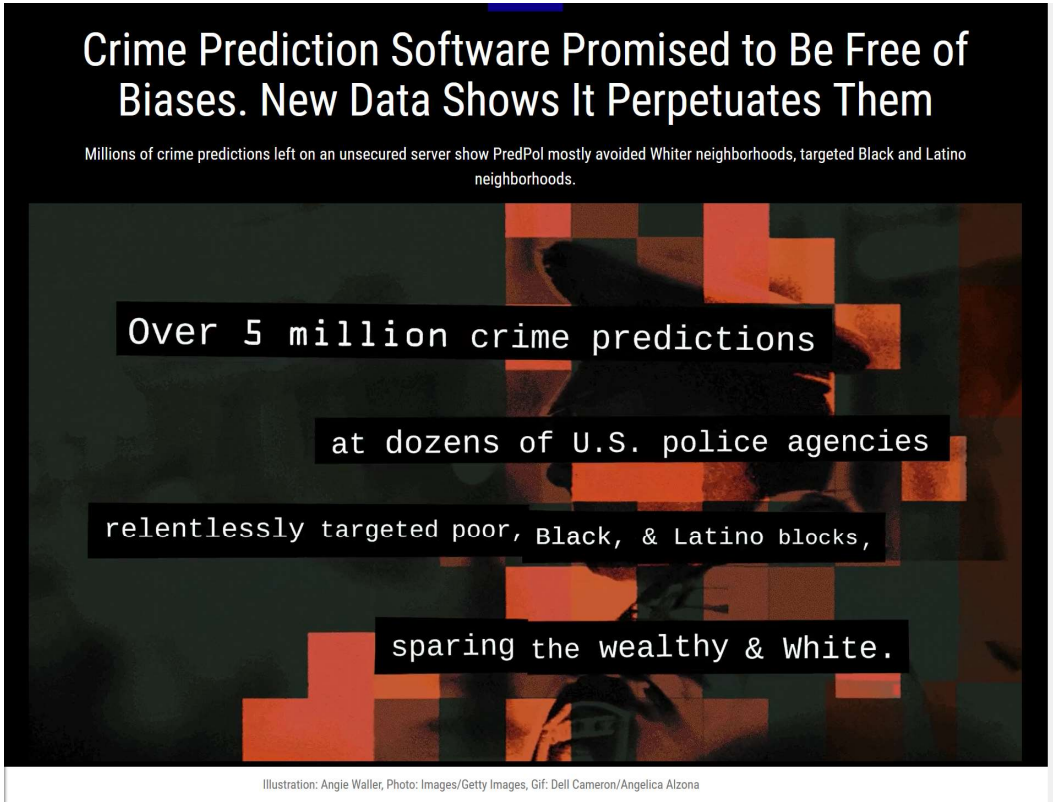
Trained on a corpus of Google News texts

The trained embedding exhibit female/male gender stereotypes, learning that "doctor" is more similar to man than to woman

- Machine learning can reproduce biases in their data
- Such learned associations accurately reflect patterns in the source text corpus
- However, their use in automated systems reproduce and amplify existing biases.

Bolukbasi, T., Chang, K-W., Zou, J.Y., Saligrama, V. and Kalai, A.T. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, 2016, 4349—4357.

Crime Prediction



Between 2018 and 2021,
more than 1 in 33 U.S. residents were potentially subject to police patrol decisions directed by crime-prediction software, PredPol.

The software recommended more patrols to neighborhoods of Blacks, Latinos, and families that would qualify for the federal free and reduced lunch program.

https://gizmodo.com/crime-prediction-software-promised-to-be-free-of-biases-1848138977

Back in 2016, neural networks start to solve difficult problems

Google Translate's App Now Instantly Translates Printed Text In 27 Languages

Posted Jul 29, 2015 by Drew Olanoff (@drew)

One of the most intense experiences you'll ever have is visiting a country that spea
language different than yours. There's a host of tools you can use, but Google's Tra

Neural translation

Google's DeepMind Masters Atari Games

Paul Rodgers, CONTRIBUTOR

I cover general science news. FULL BIO

Opinions expressed by Forbes Contributors are their own.

A computer that taught itself to play almost 50 video games including Space Invaders and Pong is being hailed as the pinnacle of artificial intelligence.

Microsoft's AI beats Ms. Pac-Man

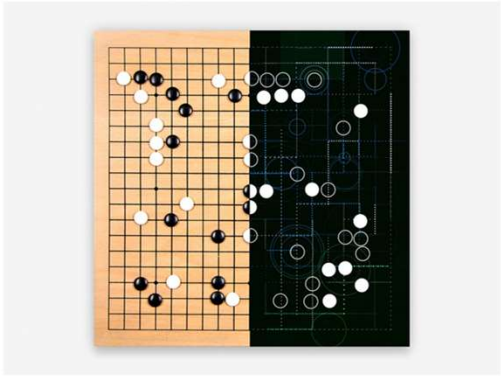
Posted Jun 15, 2017 by Brian Heater (@bheater)

Games

Deep Learning

ith massive amounts of
mputational power, machi
n now recognize objects a
anslate speech in real time.
tificial intelligence is finally
tting smart.

IN A HUGE BREAKTHROUGH, GOOGLE'S AI BEATS A TOP PLAYER AT THE GAME OF GO



GOOGLE

IN A MAJOR breakthrough for artificial intelligence, a computing system developed by Google researchers in Great Britain has beaten a top human player at the game of Go, the

obert D. Hof

hen Ray Kurzweil met with Google CEO Larr
wasn't looking for a job. A respected invent
machine-intelligence futurist, Kurzweil wa
oming book *How to Create a Mind*. He told Page, v
y draft, that he wanted to start a company to devel
to build a truly intelligent computer: one that cou
guage and then make inferences and decisions on i

Microsoft's Deep Learning Project Outperforms Humans In Image Recognition

Image Recognition

Michael Thomsen, CONTRIBUTOR

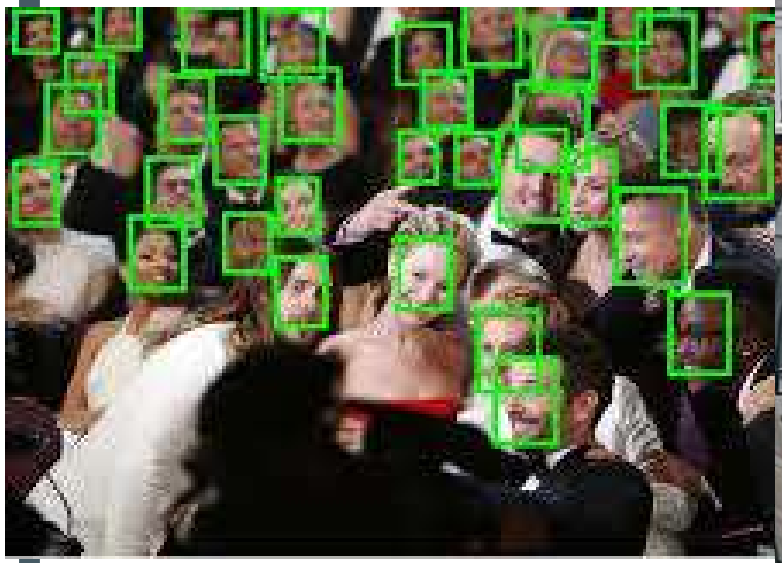
I write about tech, video games, science and culture. FULL BIO

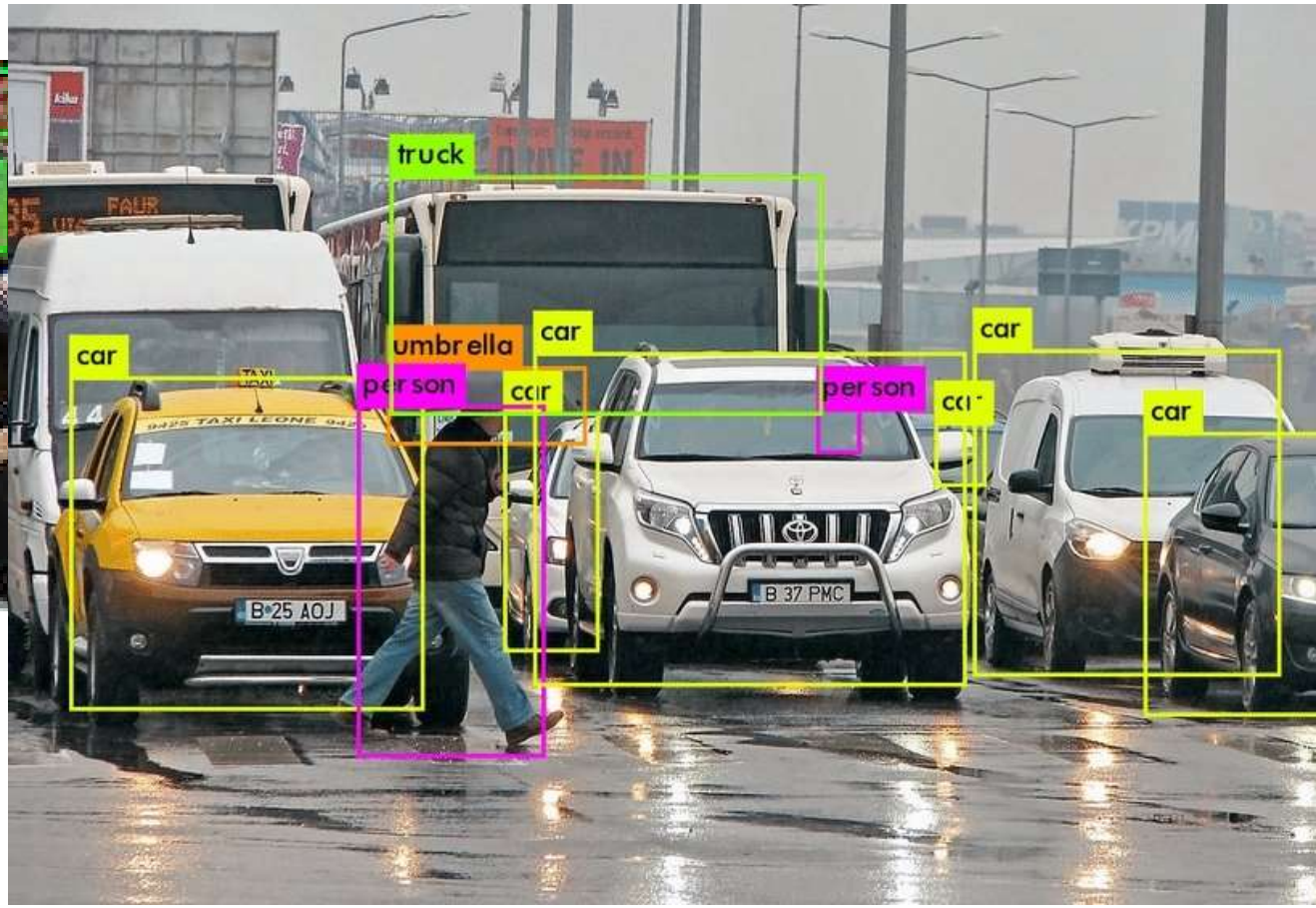
Opinions expressed by Forbes Contributors are their own.

Technology is blanketed in dishonesty. Computer phones are sma
software automations become intelligence, and coerced financialization becomes sharing. Because of the deceptive language surrounding these instruments it's difficult to talk about how they're used, and at what cost. Instead we're forced into false debates about sharing versus not sharing, intelligence versus inefficiency, progress versus everything.

Deep learning is as big a fraud as any of these endeavors, an expensive and obscure discipline built around the claim that computaters can mimic human neuronal function and thus learn as well or better than humans. This week, Microsoft MSFT +1.30% Research announced its newest deep learning project had outperformed humans in a test to identify objects in digital images. Researchers noted their scores shouldn't be taken as proof that computer image identification in general was better than humans, admitting

Object Recognition





Bias in Object Recognition

In 2020, he was accused of reaching into a vehicle, grabbing a cellphone from a man and damaging it.

Officials concluded Oliver **had been misidentified** as the perpetrator and dismissed the case.

Detroit Police used **facial recognition technology** in the investigation.

https://eu.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002/

Bias in Object Recognition

Last year, he was accused of reaching into a vehicle, grabbing a cellphone from a man and damaging it.

Officials concluded Oliver **had been misidentified** as the perpetrator

- Facial recognition systems have been used by police forces f**or more than two decades**.
- While the technology works relatively well on white men, **the results are less accurate for other demographics**
  (Recent studies by M.I.T. and the National Institute of Standards and Technology)
- In part because of a l**ack of diversity in the images** used to train the algorithm.

https://eu.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002/

Medicine

- Drug research
- Autonomous robotics labs
- Precision medicine

Today, on average, it takes more than 10 years and billions of dollars to develop a new drug.

The vision is to use AI to make drug discovery faster and cheaper.

By predicting how potential drugs might behave in the body and discarding dead-end compounds before they leave the computer, machine-learning models can cut down on the need for painstaking lab work.

https://www.technologyreview.com/2023/02/15/1067904/ai-automation-drug-development/
https://mareana.com/ai-in-accelerating-drug-development-and-production/

Bias in Medicine

If a dataset used to train an AI system lacks diversity, the AI may develop biased algorithms that perform well for certain demographic groups while failing others

- In 2023, a class action lawsuit accused UnitedHealth of illegally using an AI algorithm to turn away seriously ill elderly patients from care under Medicare Advantage.
- Another system was shown to regularly underestimate the severity of illness in Black patients because it used health care costs as a proxy for illness while failing to account for unequal access to care, and thus unequal costs, across the population.
- Female patients are disproportionately misdiagnosed for heart disease, and receive insufficient or incorrect treatment.

https://www.informationweek.com/machine-learning-ai/how-ai-bias-is-impacting-healthcare#close-modal
https://www.wired.com/story/how-algorithm-favored-whites-over-blacks-health-care/

Gen(erative) AI

AI that generates content

Examples:

Text → **Text** (Google Translate, ChatGPT, Jasper, AI-Writer, Notion AI, Lex)

Text → **Image** (Dall-E, Midjourney and Stable Diffusion)

Text → **Music** (Amper, Dadabots and MuseNet)

Text → **Code** (CodeStarter, Codex, GitHub Copilot and Tabnine)

GenAI text-to-image

A researcher presenting to other researchers

Where does bias come from?

Bias may come from:

- **the actual data**
  - if a survey contains biased questions
  - if some specific population is misrepresented in the input data
  - if the data itself is a product of a historical process that operated to the disadvantage of certain groups.

- **the algorithm**
  - reflecting, for example, commercial or other preferences of its designers (**intentional bias**)
  - data processing (**technical bias**)
  - feedback loop (**amplified bias**)

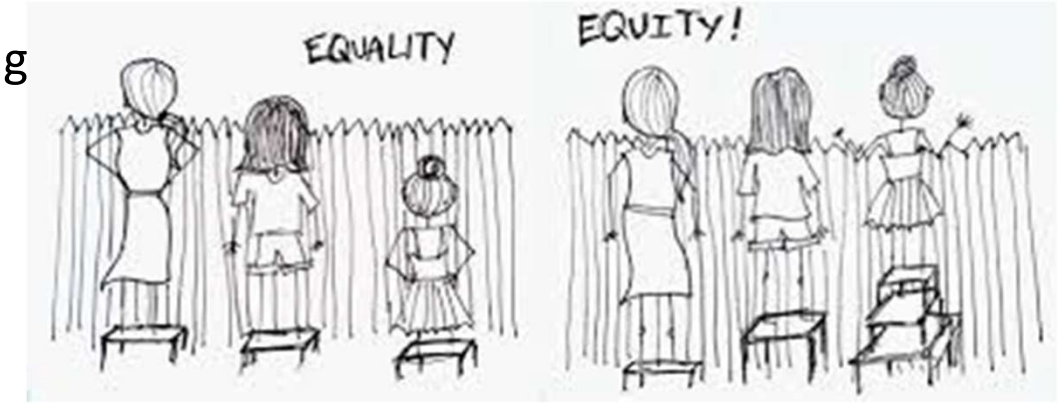
Achieving fairness (lack of discrimination)

What is a fair output?

An intricate problem from both a social and a technical perspective

From a social perspective, fundamental distinction between ***equality*** and ***equity***.

Equality: treating people equally

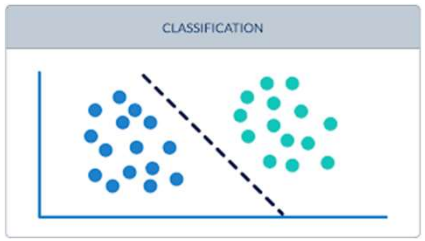


Equity: treating people according to their needs, so that they finally receive the same output

What is fair? Depends on the problem

From a technical perspective, depends on

- the ***algorithmic problem:***
  - Classification
  - Ranking
  - Recommendations
  - Clustering
  - ..



Achieving fairness (lack of discrimination)

What is a fair output?

Several fairness definitions and metrics exist:

- Based on accuracy
- Based on exposure
- etc

Fairness for:

- Users
- Items

Fairness for:

- Individuals
- Groups

Several challenges

- How do fairness definitions fare?
- Which one is suitable for which context?
- How do people perceive fairness in different contexts?

Achieving fairness

Methods for achieving algorithmic fairness can be distinguished as:

Pre-processing: Transform the data so that any underlying bias is removed

In-processing: modify existing or introduce new algorithms that produce fair outputs

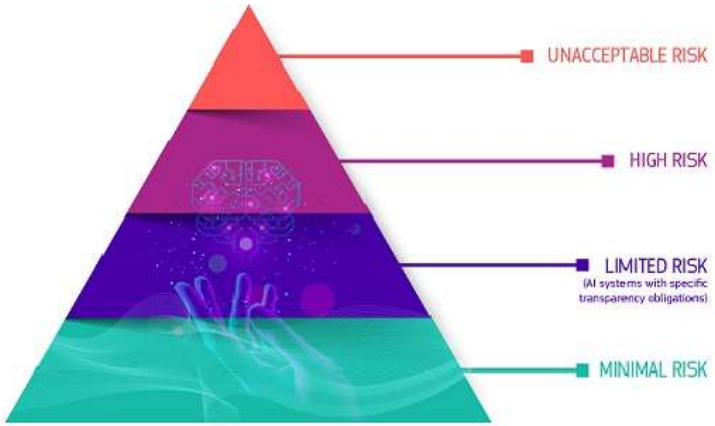
Post-processing: treat algorithms as black boxes and modify their output to ensure fairness

E. Pitoura, K. Stefanidis, G. Koutrika: Fairness in rankings and recommendations: an overview. VLDB J. 31(3): 431-458 (2022)

More is needed

EU AI Act

The AI Act aims to provide AI developers and deployers with clear requirements and obligations regarding specific uses of AI.



High risk

- critical infrastructures (e.g. transport)
- educational or vocational training
- safety components of products, etc

Before they can be put on the market:

- adequate risk assessment and mitigation systems;
- high quality of the datasets feeding the system to minimise risks and discriminatory outcomes;
- logging of activity to ensure traceability of results;
- detailed documentation to assess its compliance; etc

More is needed

200+ artists sign letter urging tech companies to stop using AI to "devalue the rights of human artists"

The Artist Rights Alliance in the US has organised an open letter signed by more than 200 artists calling on tech companies to pledge that they will only u generative AI responsibly, and that they will not infringe upon or devalue the rights of music creators

Chris Cooke

More than 200 artists - including **Nicki Minaj**, **Jonas Brothers**, **Chuck D**, **Sam Smith**, **Katy Perry** and **Zayn Malik** - have signed an open letter that calls on technology companies to "cease the use of artificial intelligence to infringe upon and devalue the rights of human artists".

Although insisting that "when used responsibly, AI has enormous potential to advance human creativity", the letter urges "AI developers, technology companies, platforms and digital music services" to "pledge that they will not develop or deploy AI music-generation technology, content or tools that undermine or replace the human artistry of songwriters and artists or deny us fair compensation for our work".

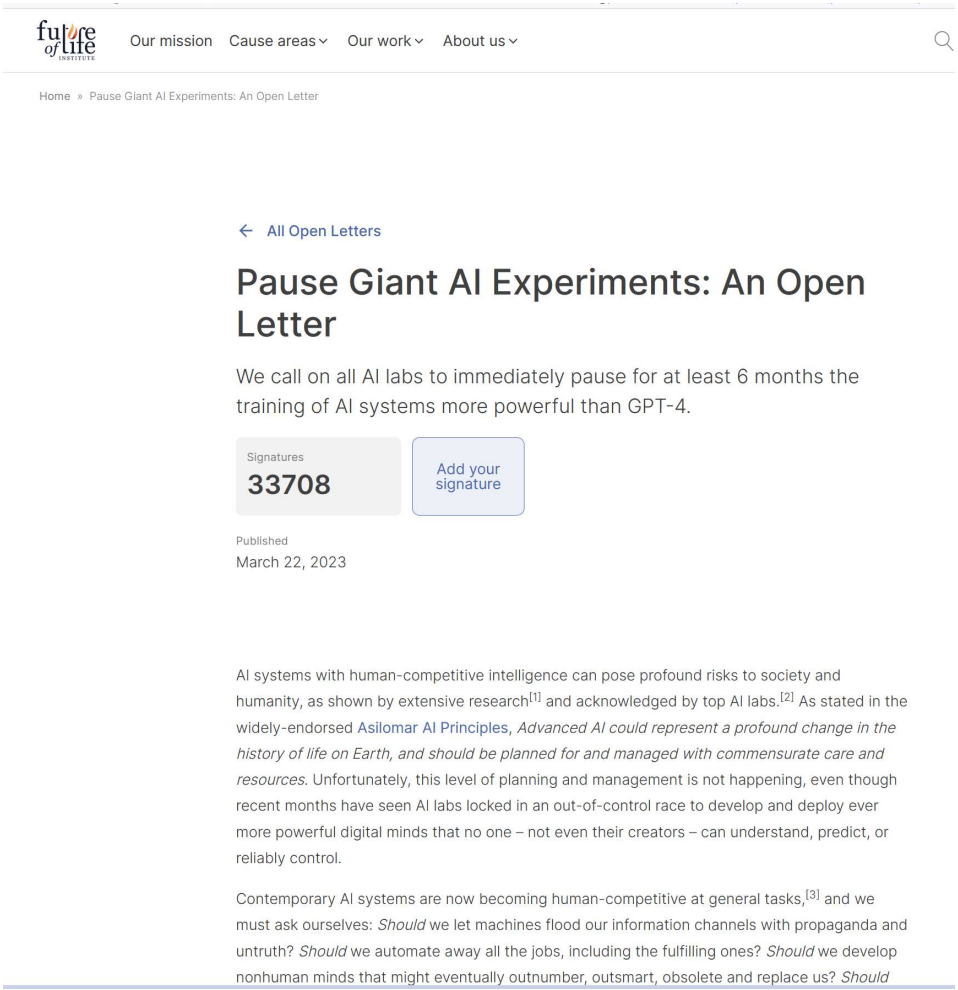
It's by no means the first such demand from the music industry, which has been clear about what it sees as the obligations of companies that are developing or

HOLLYWOOD >

In Hollywood writers' battle against AI, humans win (for now)

During the nearly five-month walkout, no issue resonated more than the use of AI in script writing. What was once a seemingly lesser demand of the Writers Guild of America became an existential rallying cry

More is needed



future of life INSTITUTE

Our mission Cause areas Our work About us

Home » Pause Giant AI Experiments: An Open Letter

← All Open Letters

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures
33708

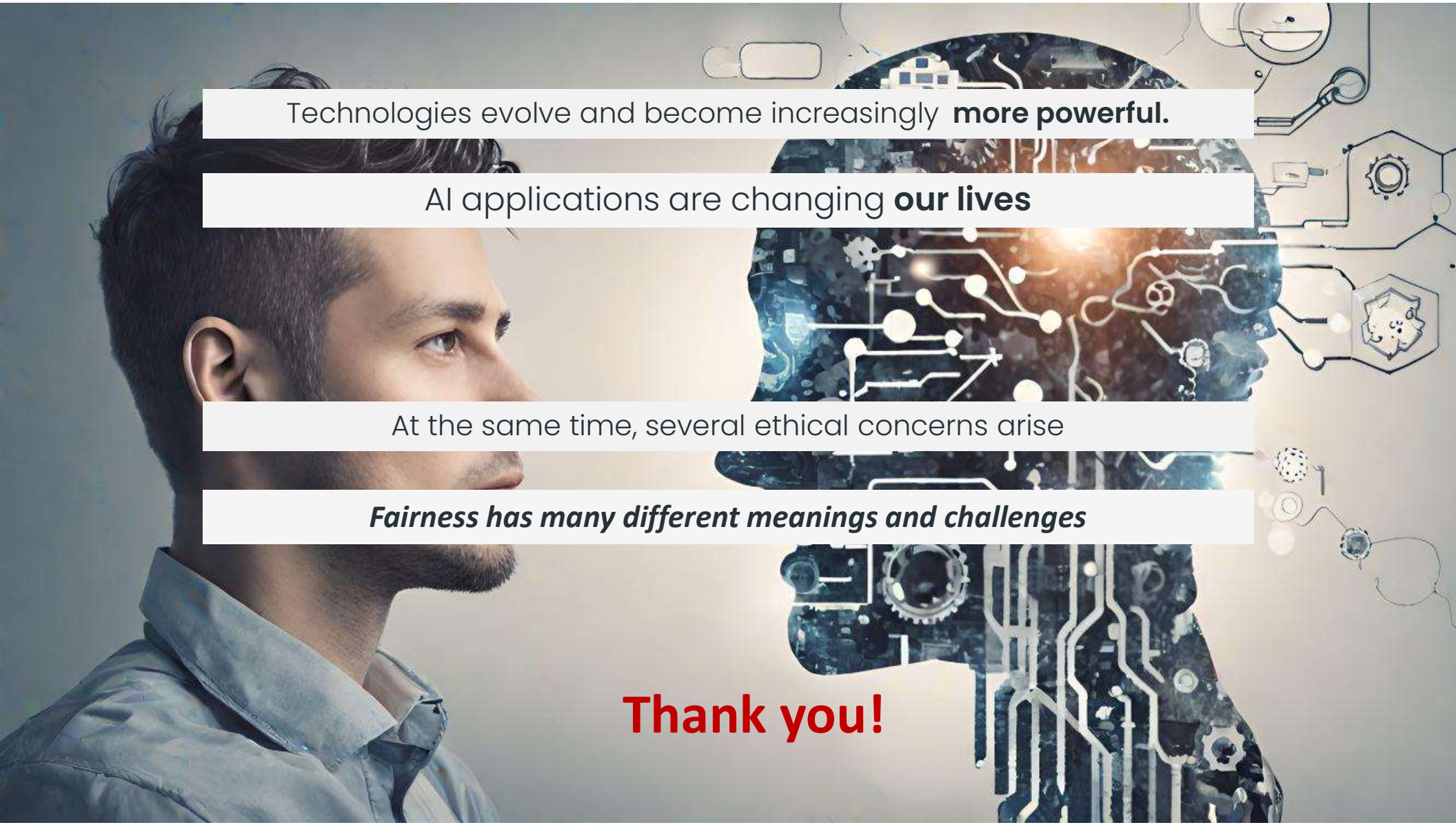
Add your signature

Published
March 22, 2023

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research[1] and acknowledged by top AI labs.[2] As stated in the widely-endorsed Asilomar AI Principles, *Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.* Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.

Contemporary AI systems are now becoming human-competitive at general tasks,[3] and we must ask ourselves: *Should* we let machines flood our information channels with propaganda and untruth? *Should* we automate away all the jobs, including the fulfilling ones? *Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? *Should*

None stopped AI experiments


Technologies evolve and become increasingly **more powerful.**
AI applications are changing **our lives**
At the same time, several ethical concerns arise
Fairness has many different meanings and challenges
Thank you!