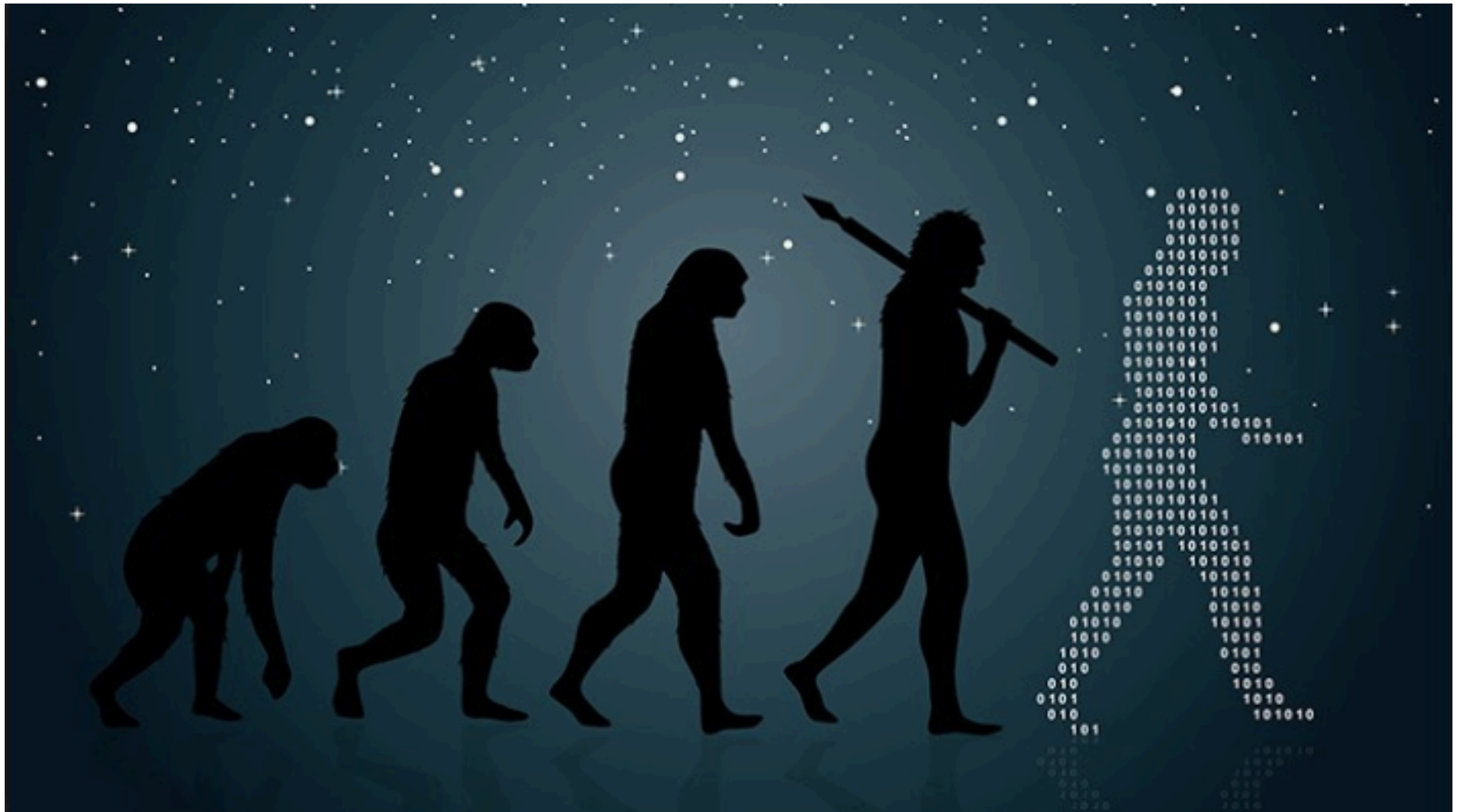**Artificial Intelligence and Ethics**

Caitlin Mulholland
caitlinsm@puc-rio.br

Big Data algorithms could create digital dictatorships in which all power is concentrated in the hands of a tiny elite while most people suffer not from exploitation but from something far worse: irrelevance.

(Yurval Harari, 21 lessons for the 21st century)

Homo Data

GENDER	HAIR COLOUR	EYES	FACE	MOUTH	BEARD	HAIR TYPE
FEMALE	BROWN	BLUE	OVAL	ROUND	FULL BEARD	CURLY
MALE	BLACK				GOATEE	WAVY
TRANSGENDER	BLOND				PENCIL	STRAIGHT
INTERSEX	RED	GREEN	SHAPE		HANDLEBAR	KINKY
NON-DEFINED	WHITE	OLIVE GREEN	TRIANGULAR		DUCKTAIL	BALD
	GREY	HONEY	HEART SHAPE		CHIN CURTAIN	DREADLOCKS

Should we use "race" as a criteria for physical description of a person?

FACIAL HAIR TYPES

Hollywoodian

Mutton Chops

A la Souvarov

French Fork

Ducktail

Fu Manchu

Old Dutch

Dali

Handlebar and Chin Puff

Van Dyke

Friendly Mutton Chops

Balbo

Rap Industry Standard

Sparrow

Klingon

Federation Standard

Short Boxed Beard

Goatee

Chin Curtain

Hulihee

The Zappa

Soul Patch

Handlebar and Goatee

The Winnfield

Petit Goatee

Franz Josef

Anchor

Napoleon III Imperial

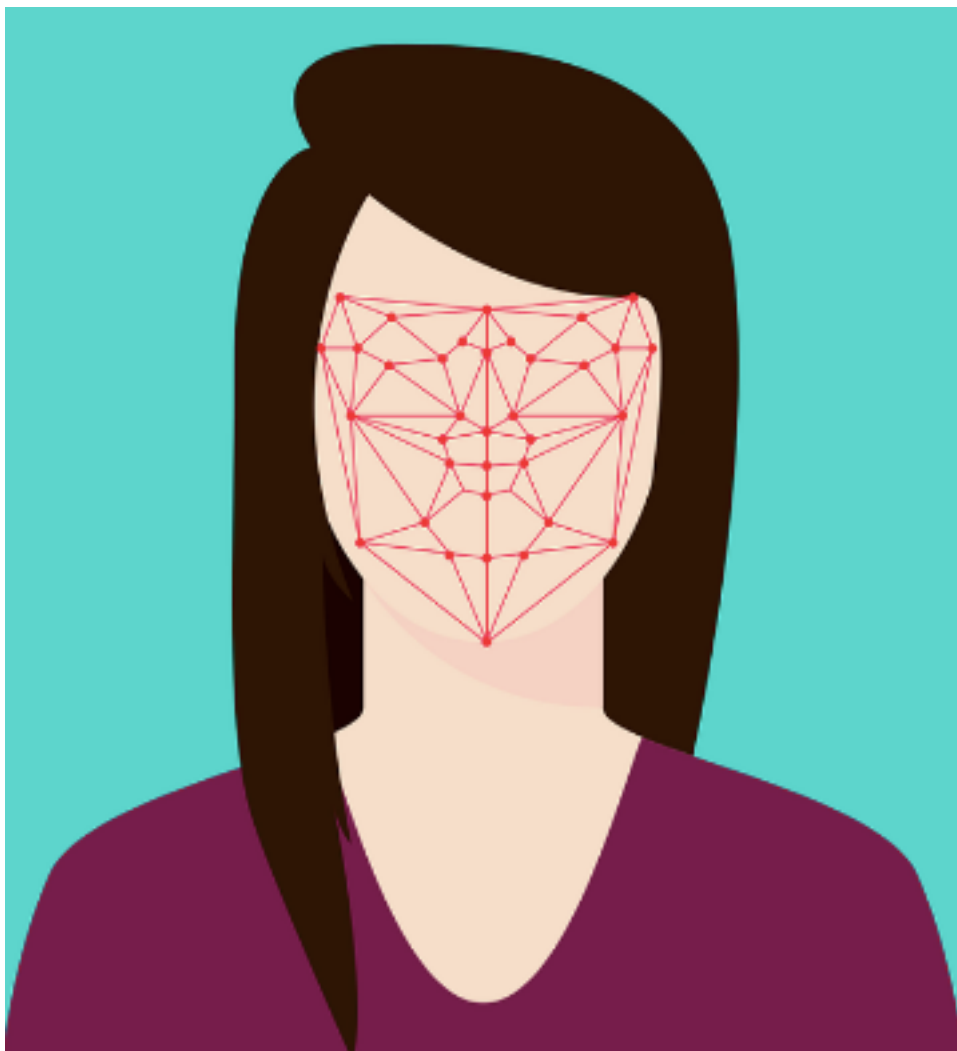Copstash Standard

Pencil

Super Mario

Handlebar

Hair Types

Iphone X and Facial Recognition

Legalite
PUC-Rio

- Unlocking Iphones using facial recognition technology.

- Does it work as a form of protecting privacy and device security?

- Well... Not in China...

Have you seen this guy?

This person does not exist

https://www.thispersondoesnotexist.com/

Produced by a GAN (generative adversarial network)
StyleGAN (Dec 2018) - Karras et al. and Nvidia
Original GAN (2014) - Goodfellow et al.
Don't panic. Learn about how it works.
Help me figure out what was learned by this AI here.
Check out text generation by another AI
Click for another person Link to image

AI - What you "put in" is what you "get out"
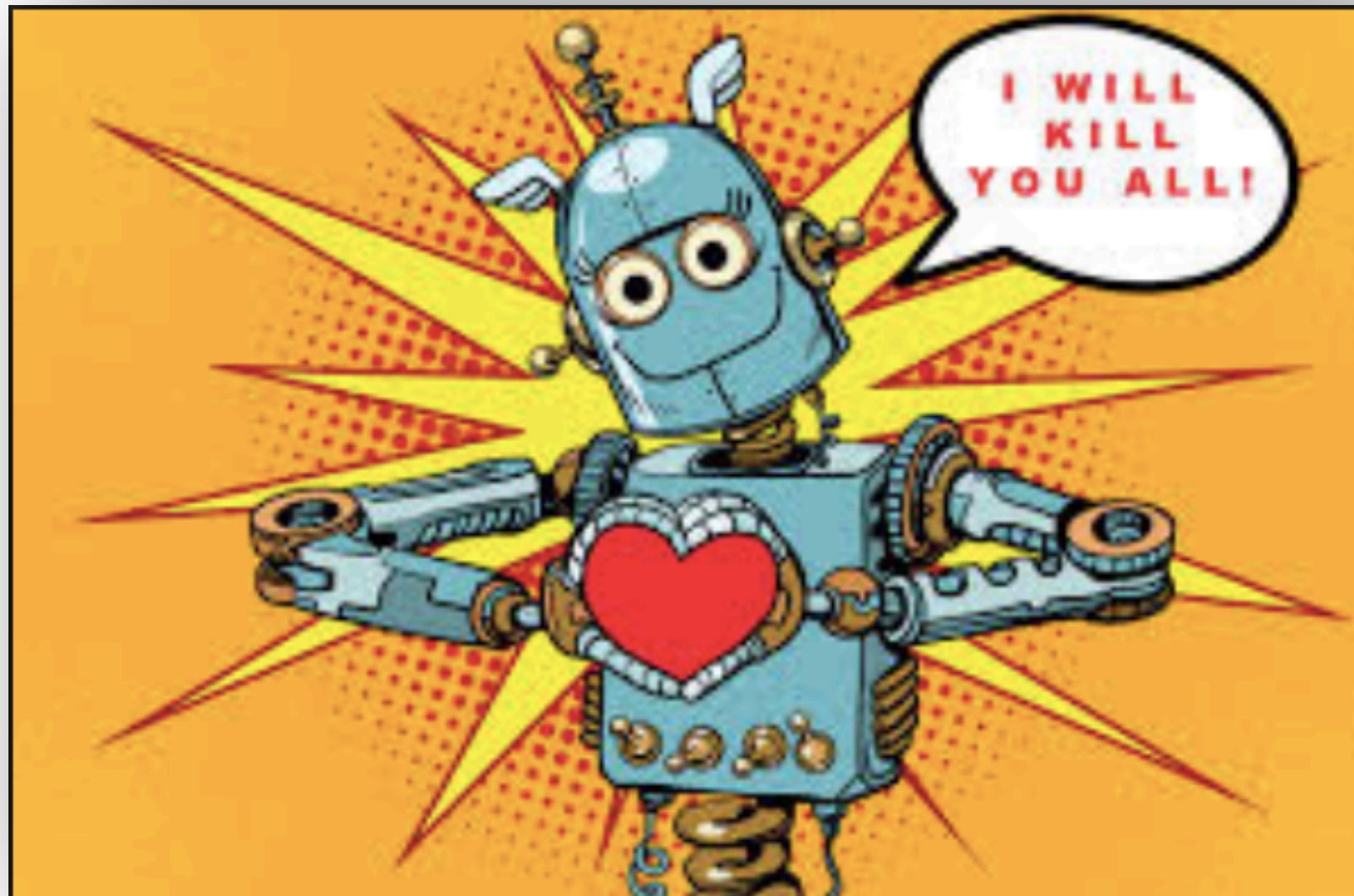
AI's white guy problem isn't going away

Legalite
PUC-Rio

- Women -
18% of authors at leading AI conferences,
20% of AI professorships, and
15% of research staff at Facebook and
10% of research staff at Google.

- Black workers -
2.5% of Google's entire workforce and
4% of Facebook's and Microsoft's.

- Transgender people - ?

- Other gender minorities - ?

(Reference: Karen Hao, MIT Tech Review, Apr 17, 2019)

Female under-representation in ICTs

- Machines assume values and function as dictated by their creators.

- Different results require different lived experiences, priorities and worldviews.

- Biases

- Prejudices

- Lack of empathy
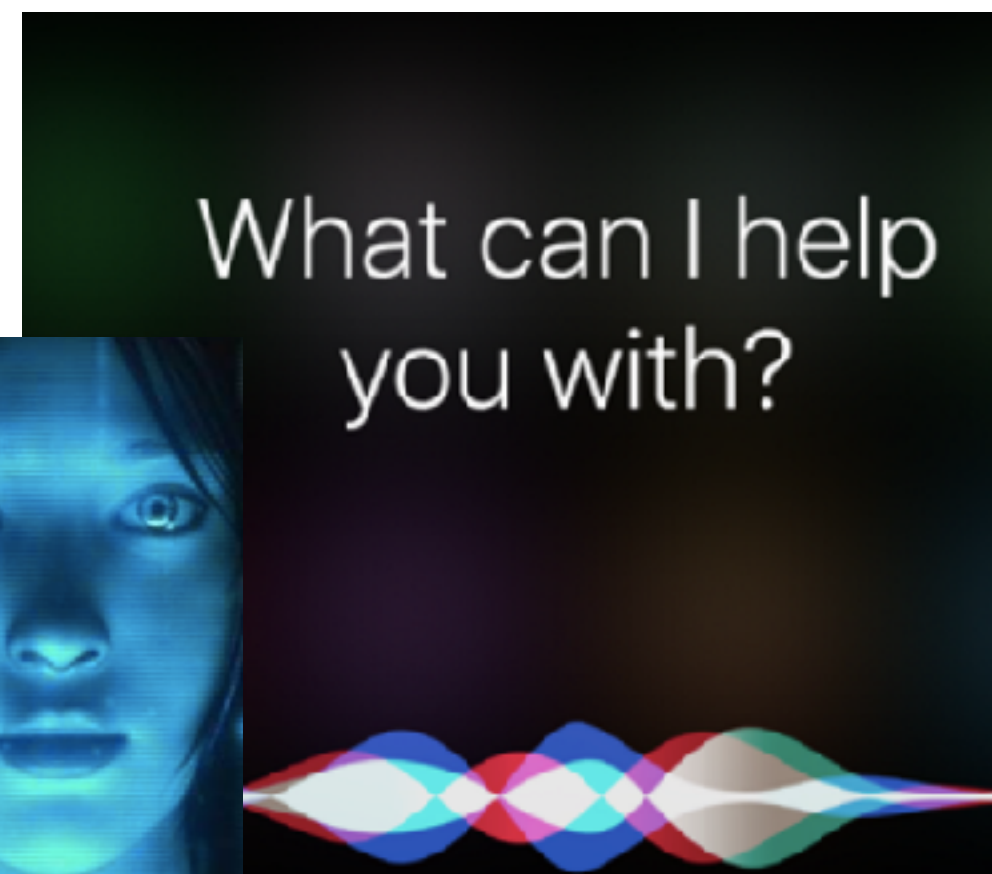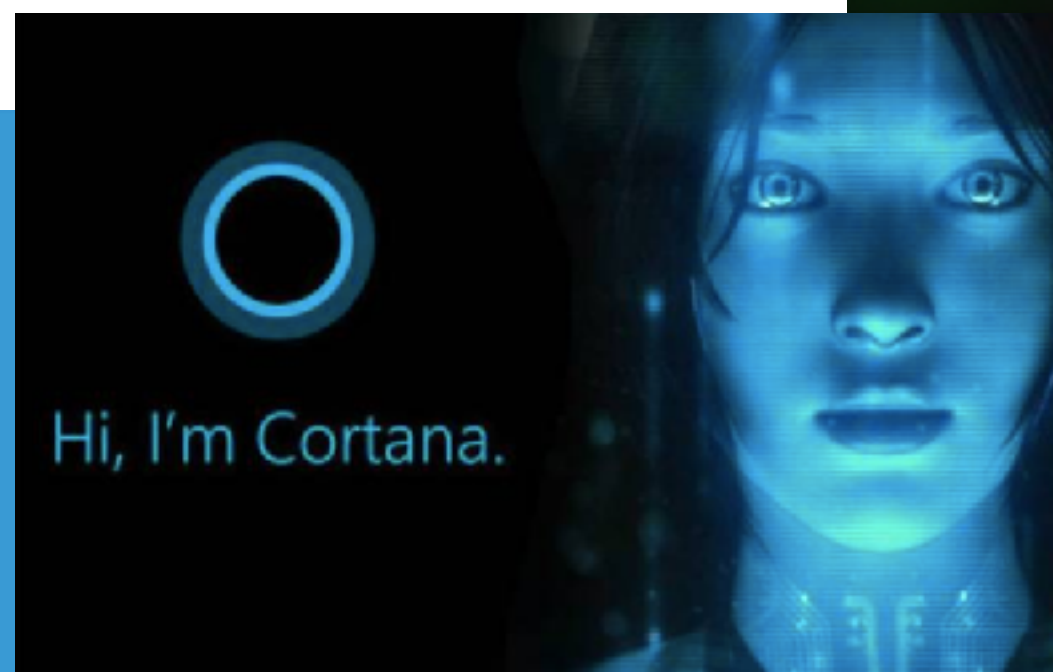
- Possible statistical misleadings

What to do about the "biases"

- Ensure that programmers are recruited from diverse backgrounds of gender, ethnicity, and socioeconomic status.

    - And why is that?

- IA looks at available real-world data and applies it to its algorithms.

Gender stereotypes in AI

Voice of female AI - helpful, subservient, and non-threatening women.

he is a soldier
she's a teacher
he is a doctor
she is a nurse
he is a writer
she is a nanny
he is a president
he is an entrepreneur
she is a singer
he is hard working
she is lazy
he is an engineer

he is an architect
he is a dentist
he is a lawyer
she is beautiful
he is very beautiful

Different names, different races…

20 "Whitest" Girl Names

Molly

Amy

Claire

Emily

Madeline

Emma

Abigail

Carly

Jenna

Heather

Katherine

Caitlin

Holly

Allison

Hannah

20 "Blackest" Girl Names

Imani

Ebony

Shanice

Aaliyah

Precious

Asia

Aliyah

Jada

Tierra

Tiara

Kiara

Jazmine

Can AI use name patterns to grant loans for housing purposes?

What is Q?

- https://www.genderlessvoice.com/

Who made Q?
Copenhagen Pride, Virtue, Equal AI, Koalition Interactive & thirtysoundsgood.

Why did we make Q?
Technology companies often choose to gender technology believing it will make people more comfortable adopting it.

Unfortunately this reinforces a binary perception of gender, and perpetuates stereotypes that many have fought hard to progress.

As society continues to break down the gender binary, recognising those who neither identify as male nor female, the technology we create should follow.

Q is an example of what we hope the future holds; a future of ideas, inclusion, positions and diverse representation in technology.

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs
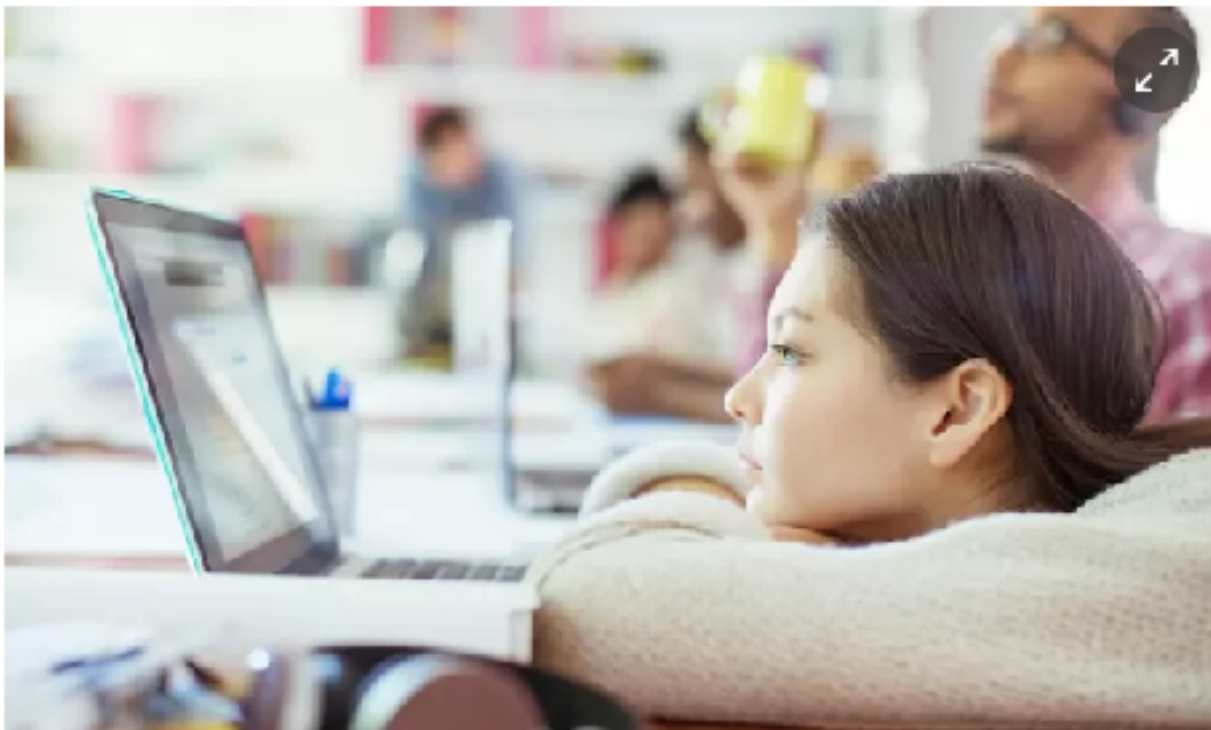
▲ One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

Highly paid jobs were not shown as often for women's and men's surveys because of the way their algorithms were written.

Early users of the site job search function were predominantly male for those high-paying jobs, so they ended up offering those jobs to men - simply reinforcing the bias against women.
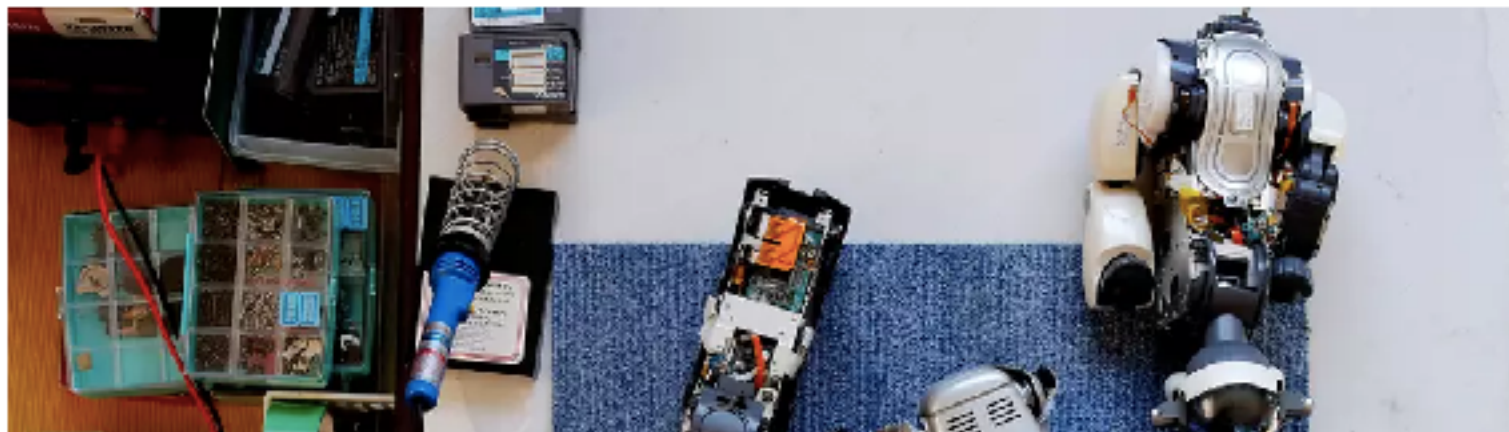
Voices

This is how artificial intelligence is undoing women's rights

Through AI we are unpicking the seams of what feminism and the women's movement has long fought for: agency, women's liberation and the idea that men and women have equal capabilities

Dejan Jotanovic Tuesday 19 June 2018 15:11 | 8 comments

Click to follow
The Independent Voices

the underlying biases of these black boxes need to be understood.

"We take the biases, which in certain ways is what we call 'culture', put it in a black box and crystallize it forever. That's where we have a problem. We have even more problems when we think the black box contains the truth and we follow it blindly, and we say, "The computer says no." That's the problem"

AI and Ethics

- The AI should reflect the diversity of users it serves.

- The AI should be held accountable - and users too (accountability)

- Reward the AI for "showing how it works" (training)

- The AI should level the field of action (accessibility)

- The AI will replace people, but it must also create new opportunities.

There are three main steps for the human rights due diligence process:

i. Identify possible discriminatory results

ii. Take effective measures to prevent and mitigate discrimination and track responses

iii. Be transparent about efforts to identify, prevent and mitigate discrimination in machine learning systems.

After identifying the risks to human rights, the second step is to avoid these risks. For developers of machine learning systems, this requires:

a) Correct the discrimination, both in the design of the model and in the impact of the system, and in deciding what training data to use.

b) Seek diversity, equity, and other means of inclusion in the machine learning development teams, in order to identify bias by design and prevent inadvertent discrimination.

c) Submit systems that have a significant risk of resulting in human rights violations to third party independent audits.

EU AI Ethical Guidelines

- Human agency and oversight

- Technical robustness and safety

- Privacy and data governance

- Transparency

- Diversity, non-discrimination, and fairness

- Environmental and societal well-being

- Accountability

Change AI to change the world

Conheça mais sobre o núcleo, suas pesquisas e projetos em
www.puc-rio.br/legalite

Caitlin Mulholland

caitlinsm@puc-rio.br

Legalite
PUC-Rio