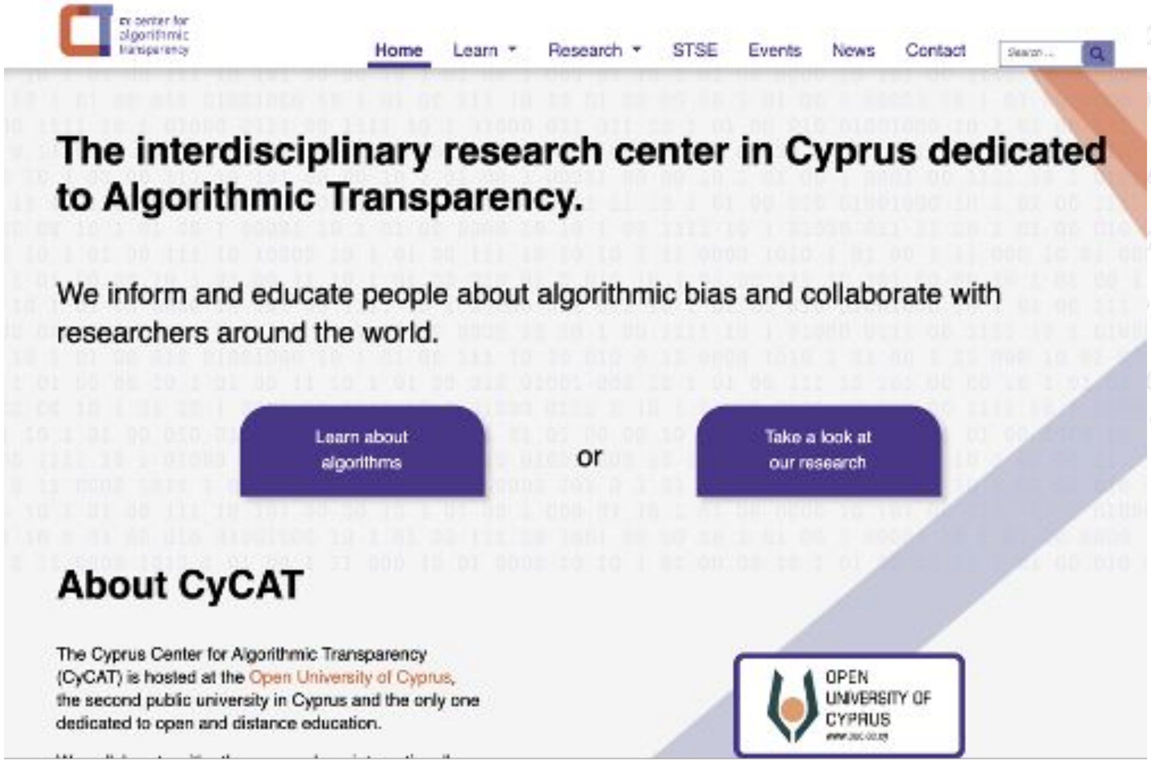
FATE: Fairness, Accountability, Transparency and Ethics

An introduction for developers

Styliani Kleanthous Loizou, Ph.D.
Kalia Orphanou, Ph.D.
Jahna Otterbacher, Ph.D.

CY. center for algorithmic transparency

ΑΝΟΙΚΤΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΚΥΠΡΟΥ
www.ouc.ac.cy


cy center for algorithmic transparency
Home
Learn ▾
Research ▾
STSE
Events
News
Contact
Search ...
Learn about algorithms
or
Take a look at our research
OPEN UNIVERSITY OF CYPRUS
www.ouc.ac.cy

The interdisciplinary research center in Cyprus dedicated to Algorithmic Transparency.

We inform and educate people about algorithmic bias and collaborate with researchers around the world.

About CyCAT

The Cyprus Center for Algorithmic Transparency (CyCAT) is hosted at the Open University of Cyprus, the second public university in Cyprus and the only one dedicated to open and distance education.

אוניברסיטת חיפה
University of Haifa
جامعة حيفا

UNIVERSITY OF TRENTO - Italy

THE UNIVERSITY of EDINBURGH

The University Of Sheffield.

Interventions for Awareness

- Three audiences:
  - Public school teachers (*in collaboration with the Cyprus Pedagogical Institute*)
    - Three-hour seminar and evaluation
  - Developers (*in collaboration with UCY - CS*)
    - 10-hour seminar and evaluation
  - General public
    - Tool-based intervention

Developer Seminar Objectives

In this 10-hour seminar participants will:

- Become aware of FATE issues in the development of (algorithmic) process/systems
- Learn core FATE concepts related to software development
- Develop appreciation for the role that developers play in mitigating algorithmic bias and in promoting ethical practices
- Experiment for techniques for auditing services / modules used in development

Overview - Day 1

Pre-seminar Questionnaire	14.10 - 14.40
Introduction to FATE	14.40 - 15.45
Break	15.45 - 16.00
FATE as a scientific field	16.00 - 17.00
Exercise in breakout rooms	17.00 - 17.30
Discussion and final thoughts	17.30 - 18.00

Overview - Day 2

Overview and questions	14.00 - 14.10
COMPAS case study discussion	14.10 - 14.40
FATE Problems	14.40 - 15.10
Break	15.10 - 15.25
FATE Solutions	15.25 - 16.25
Exercise in breakout rooms	16.25 - 17.00
Post-seminar questionnaire	17.00 - 17.30
Discussion and final thoughts	17.30 - 18.00

PRE-SEMINAR QUESTIONNAIRE

https://forms.gle/KiuNQACwZRMNh8H36

Introduction to FATE

Fairness, Accountability, Transparency and Ethics

CY. center for
algorithmic
transparency

AI and Industrial Development

Three priorities:

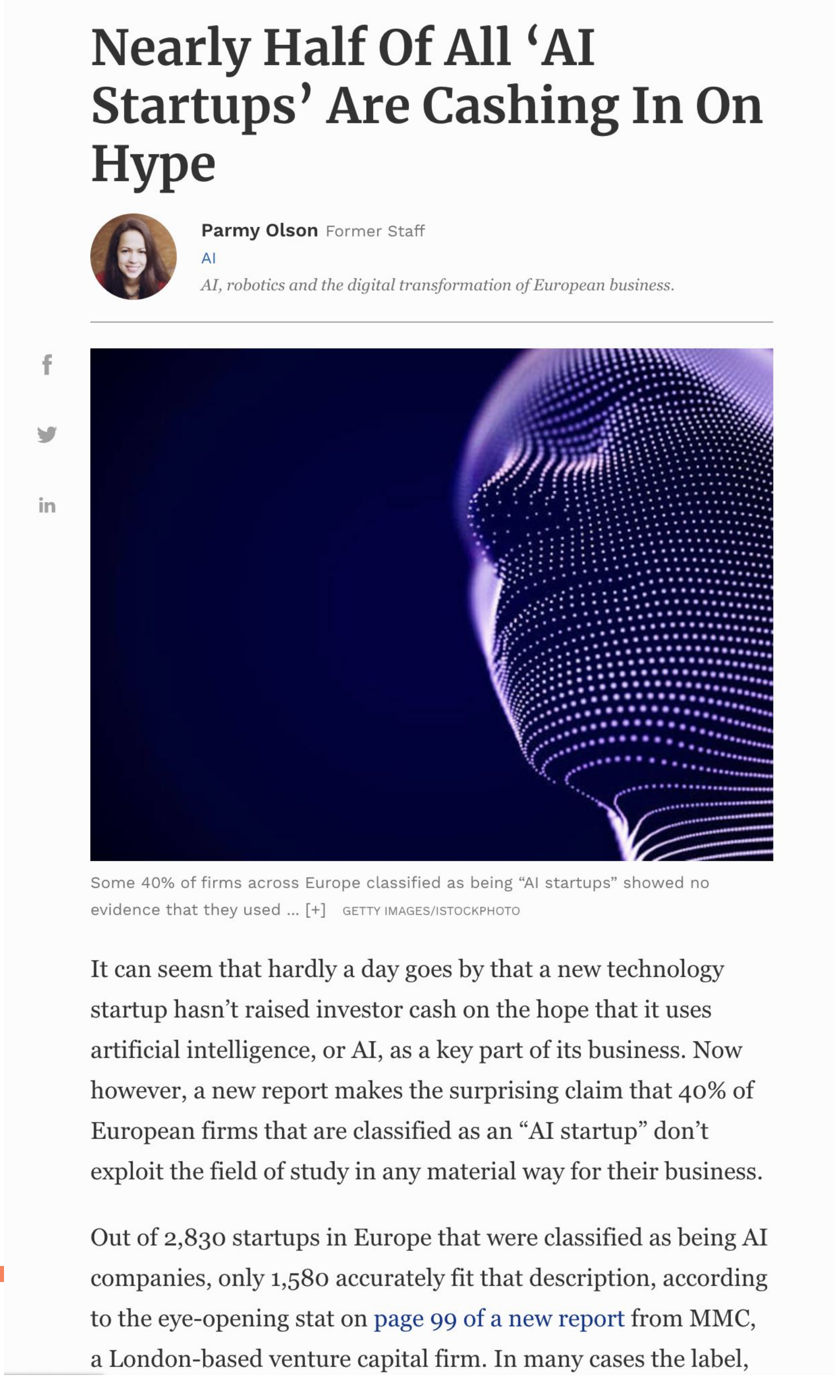
- Manufacturing - IoTs
- Mobility
- Smart Health

European Commission

Artificial intelligence – critical industrial applications

Report on market analysis of prioritised value chains, the most critical AI applications and the conditions for AI rollout

December 2019



Nearly Half Of All 'AI Startups' Are Cashing In On Hype

Parmy Olson Former Staff

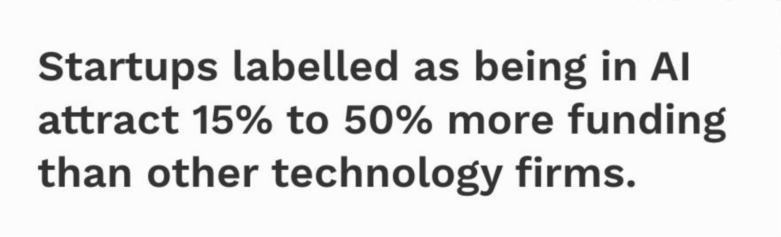
AI

AI, robotics and the digital transformation of European business.

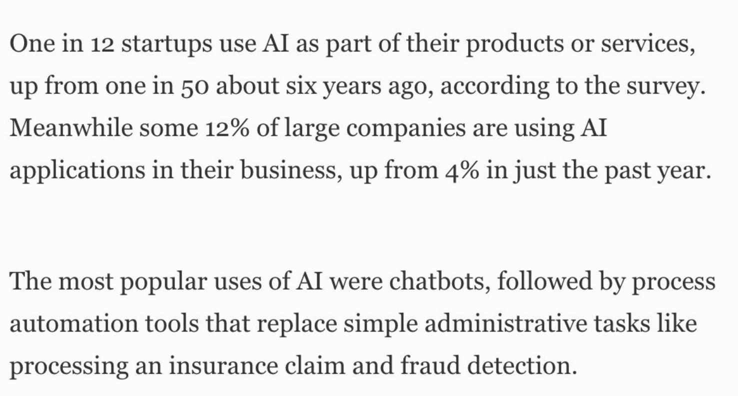
Some 40% of firms across Europe classified as being "AI startups" showed no evidence that they used ... [+] GETTY IMAGES/ISTOCKPHOTO

It can seem that hardly a day goes by that a new technology startup hasn't raised investor cash on the hope that it uses artificial intelligence, or AI, as a key part of its business. Now however, a new report makes the surprising claim that 40% of European firms that are classified as an "AI startup" don't exploit the field of study in any material way for their business.

Out of 2,830 startups in Europe that were classified as being AI companies, only 1,580 accurately fit that description, according to the eye-opening stat on page 99 of a new report from MMC, a London-based venture capital firm. In many cases the label,



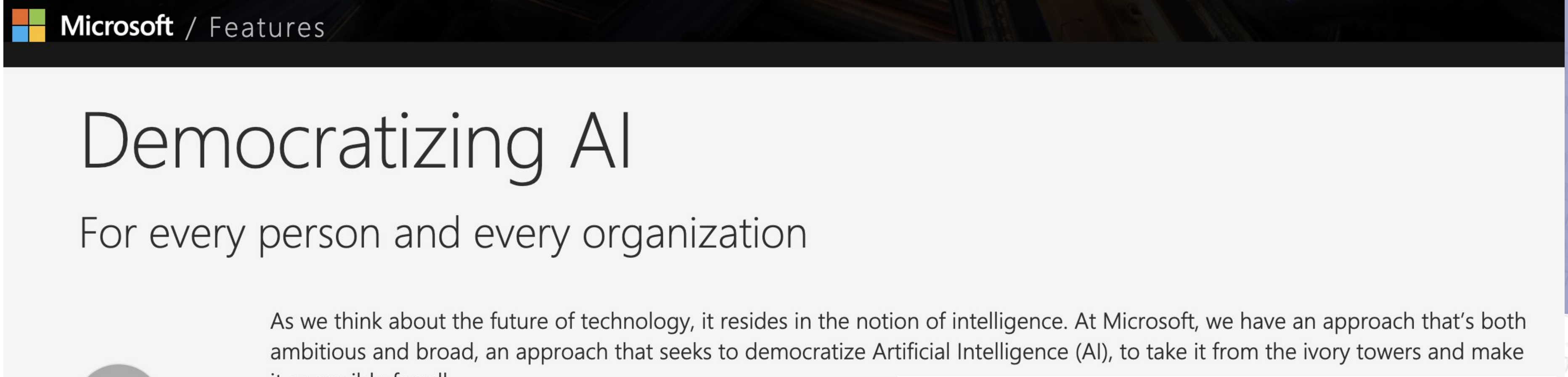
Startups labelled as being in AI attract 15% to 50% more funding than other technology firms.



One in 12 startups use AI as part of their products or services, up from one in 50 about six years ago, according to the survey. Meanwhile some 12% of large companies are using AI applications in their business, up from 4% in just the past year.

The most popular uses of AI were chatbots, followed by process automation tools that replace simple administrative tasks like processing an insurance claim and fraud detection.

https://www.forbes.com/sites/parmyolson/2019/03/04/nearly-half-of-all-ai-startups-are-cashing-in-on-hype/



Microsoft / Features

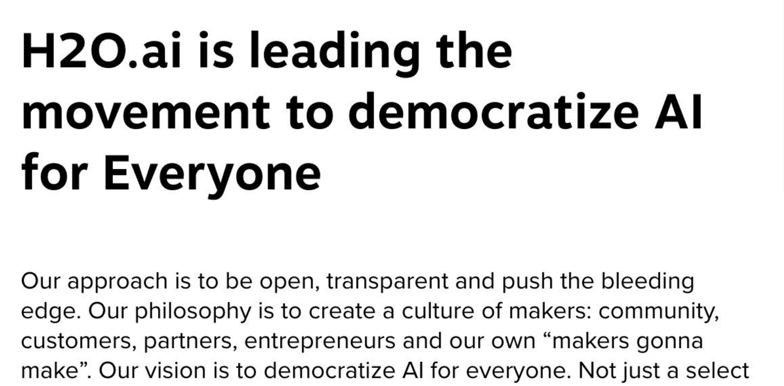
Democratizing AI

For every person and every organization

Microsoft
News Center

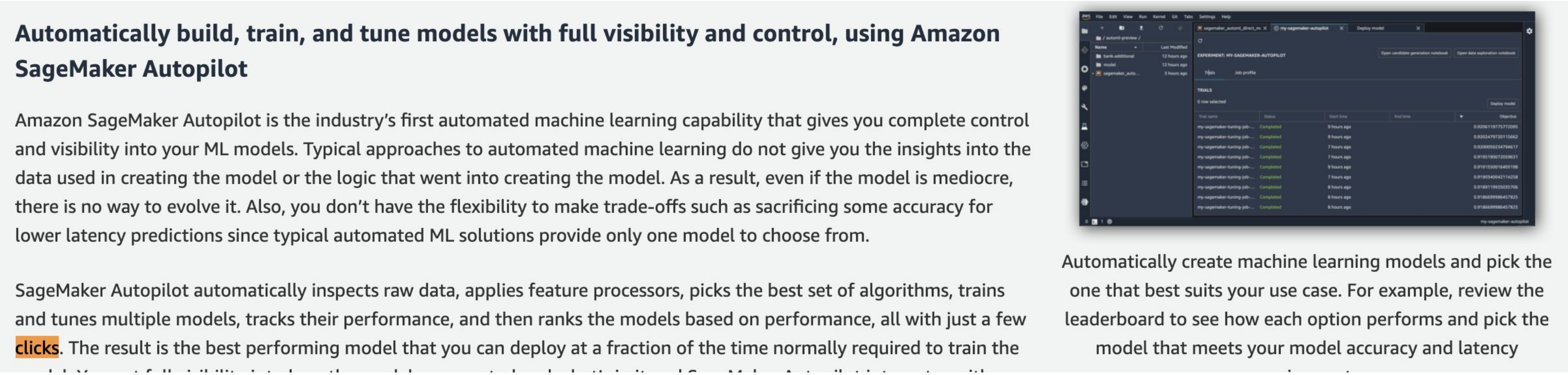
As we think about the future of technology, it resides in the notion of intelligence. At Microsoft, we have an approach that's both ambitious and broad, an approach that seeks to democratize Artificial Intelligence (AI), to take it from the ivory towers and make it accessible for all.

And as we consider the future, it's often instructive to look to the information. With the advent of the printing press in the 1400s w event around access that made it possible for humans everywhe



H2O.ai is leading the movement to democratize AI for Everyone

Our approach is to be open, transparent and push the bleeding edge. Our philosophy is to create a culture of makers: community, customers, partners, entrepreneurs and our own "makers gonna make". Our vision is to democratize AI for everyone. Not just a select



Automatically build, train, and tune models with full visibility and control, using Amazon SageMaker Autopilot

Amazon SageMaker Autopilot is the industry's first automated machine learning capability that gives you complete control and visibility into your ML models. Typical approaches to automated machine learning do not give you the insights into the data used in creating the model or the logic that went into creating the model. As a result, even if the model is mediocre, there is no way to evolve it. Also, you don't have the flexibility to make trade-offs such as sacrificing some accuracy for lower latency predictions since typical automated ML solutions provide only one model to choose from.

SageMaker Autopilot automatically inspects raw data, applies feature processors, picks the best set of algorithms, trains and tunes multiple models, tracks their performance, and then ranks the models based on performance, all with just a few clicks. The result is the best performing model that you can deploy at a fraction of the time normally required to train the

Automatically create machine learning models and pick the one that best suits your use case. For example, review the leaderboard to see how each option performs and pick the model that meets your model accuracy and latency

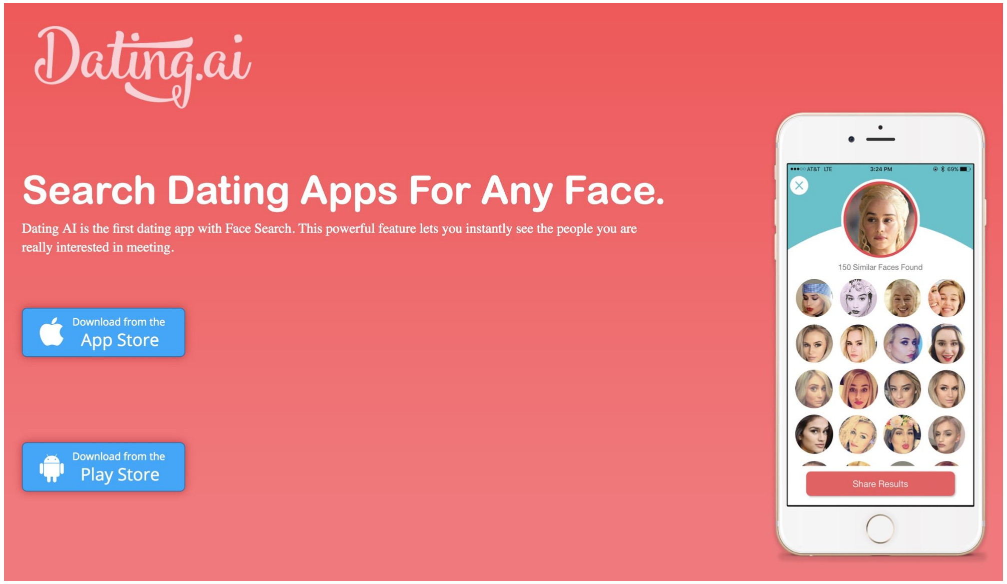
Microsoft
Face API

Google Cloud
Vision API

IBM
Watson™

clarifai

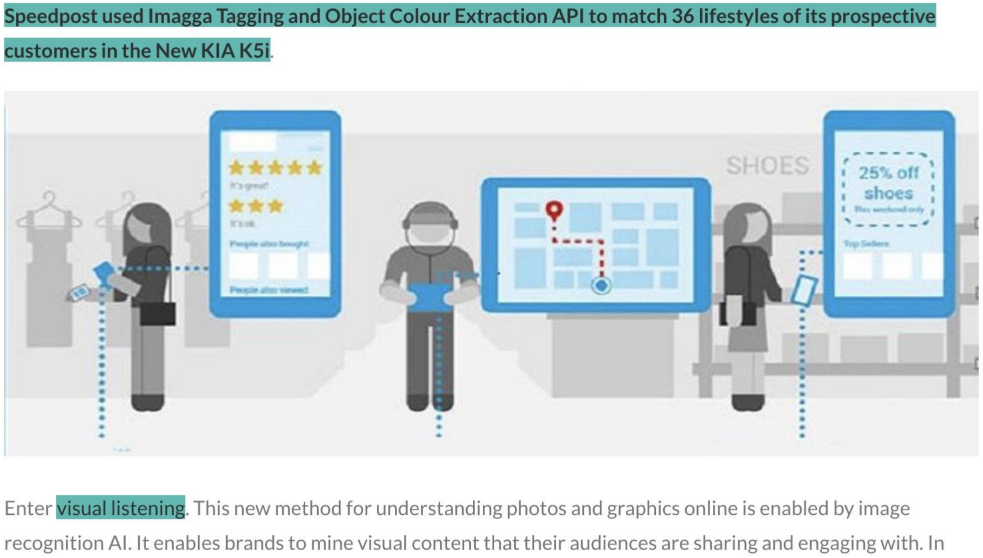
amazon
web services
Rekognition



Dating.ai

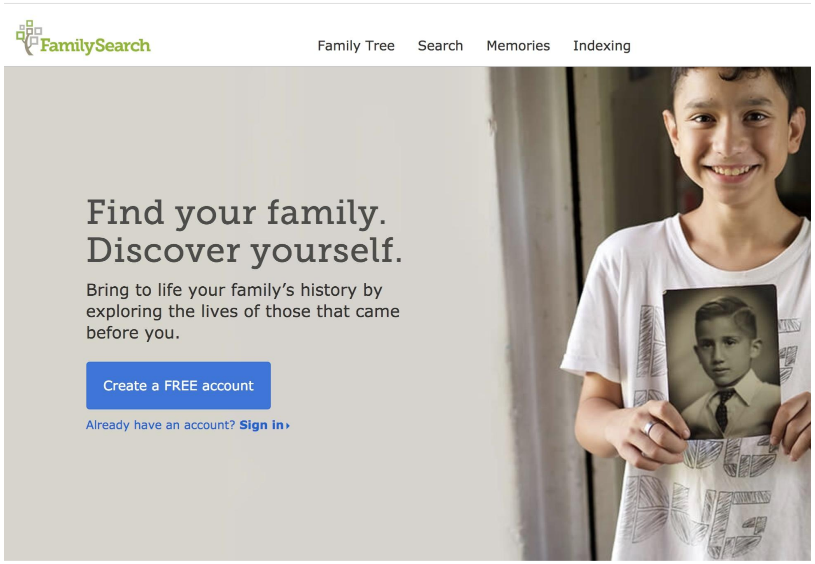
Search Dating Apps For Any Face.

Dating AI is the first dating app with Face Search. This powerful feature lets you instantly see the people you are really interested in meeting.



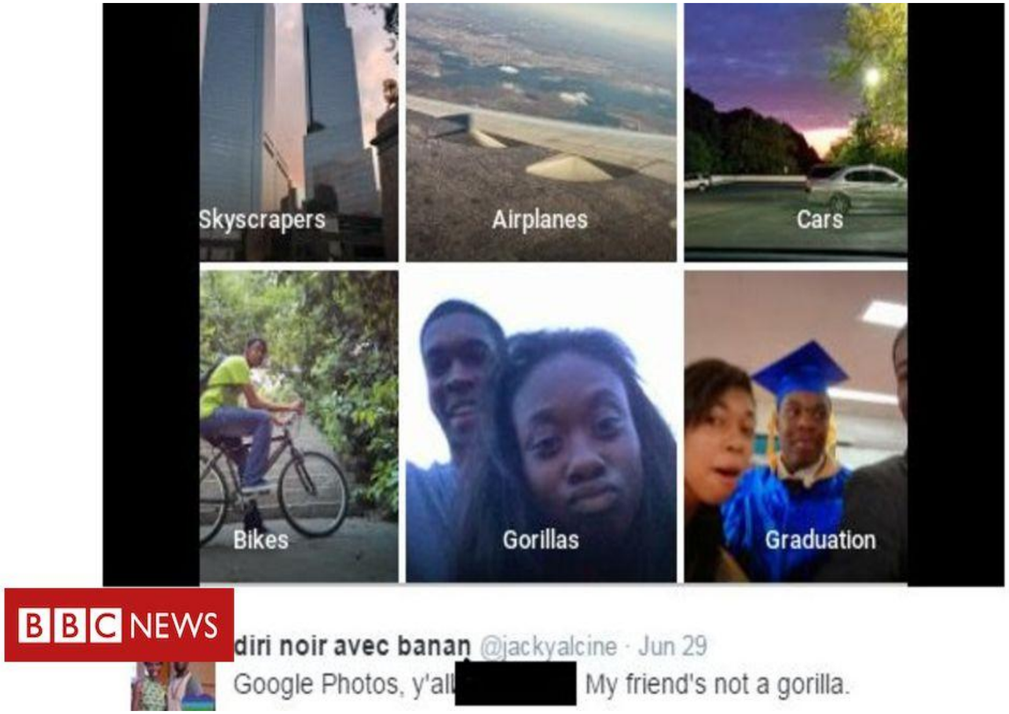
Speedpost used Imagga Tagging and Object Colour Extraction API to match 36 lifestyles of its prospective customers in the New KIA K5i.

Enter visual listening. This new method for understanding photos and graphics online is enabled by image recognition AI. It enables brands to mine visual content that their audiences are sharing and engaging with. In



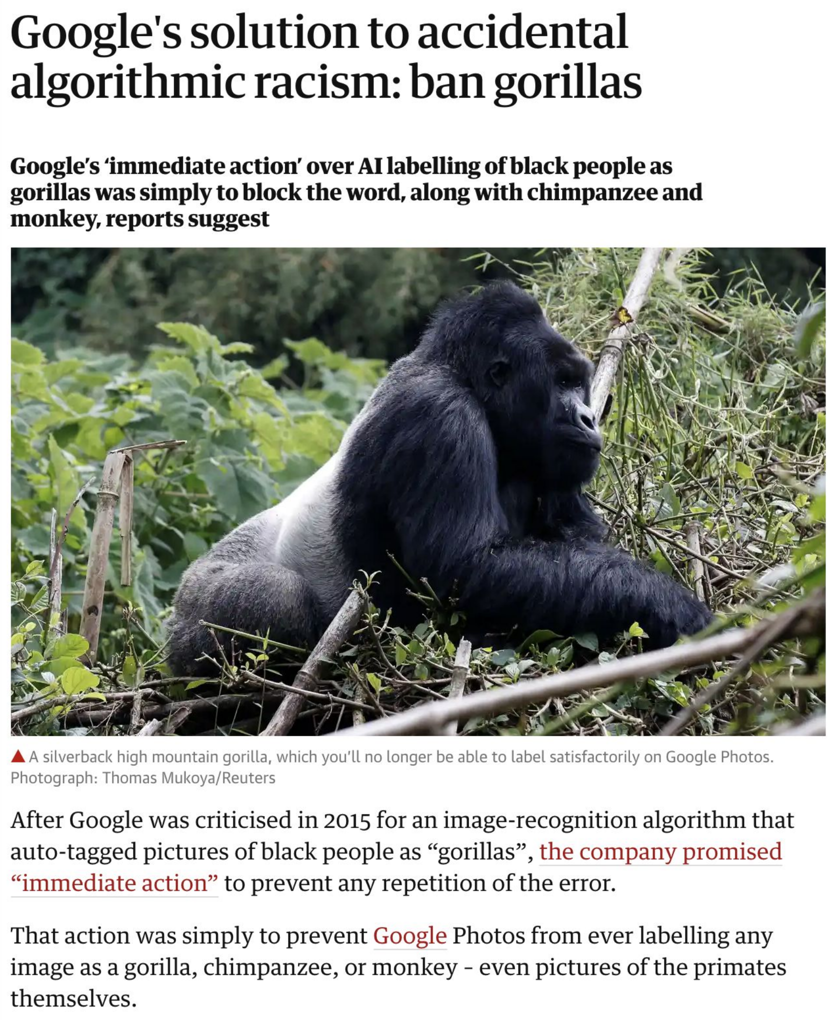
Find your family.
Discover yourself.

Bring to life your family's history by exploring the lives of those that came before you.



diri noir avec banan @jackyalcine · Jun 29

Google Photos, y'all [redacted] My friend's not a gorilla.



Google's solution to accidental algorithmic racism: ban gorillas

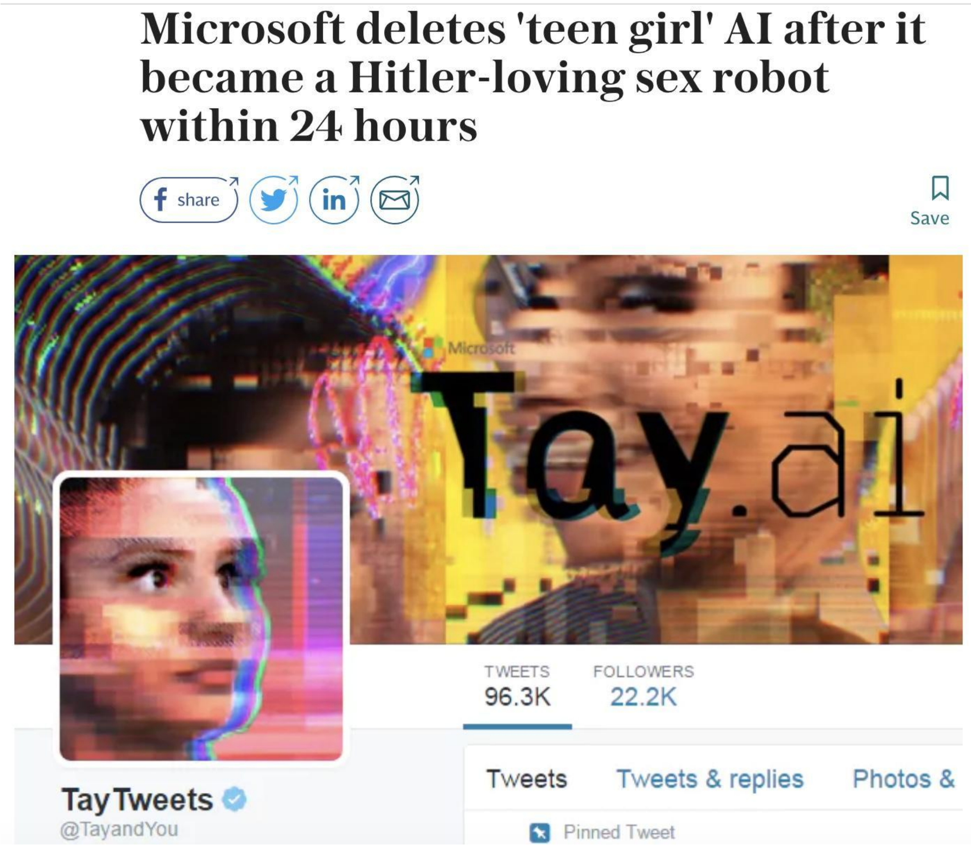
Google's 'immediate action' over AI labelling of black people as gorillas was simply to block the word, along with chimpanzee and monkey, reports suggest

▲ A silverback high mountain gorilla, which you'll no longer be able to label satisfactorily on Google Photos. Photograph: Thomas Mukoya/Reuters

After Google was criticised in 2015 for an image-recognition algorithm that auto-tagged pictures of black people as "gorillas", the company promised "immediate action" to prevent any repetition of the error.

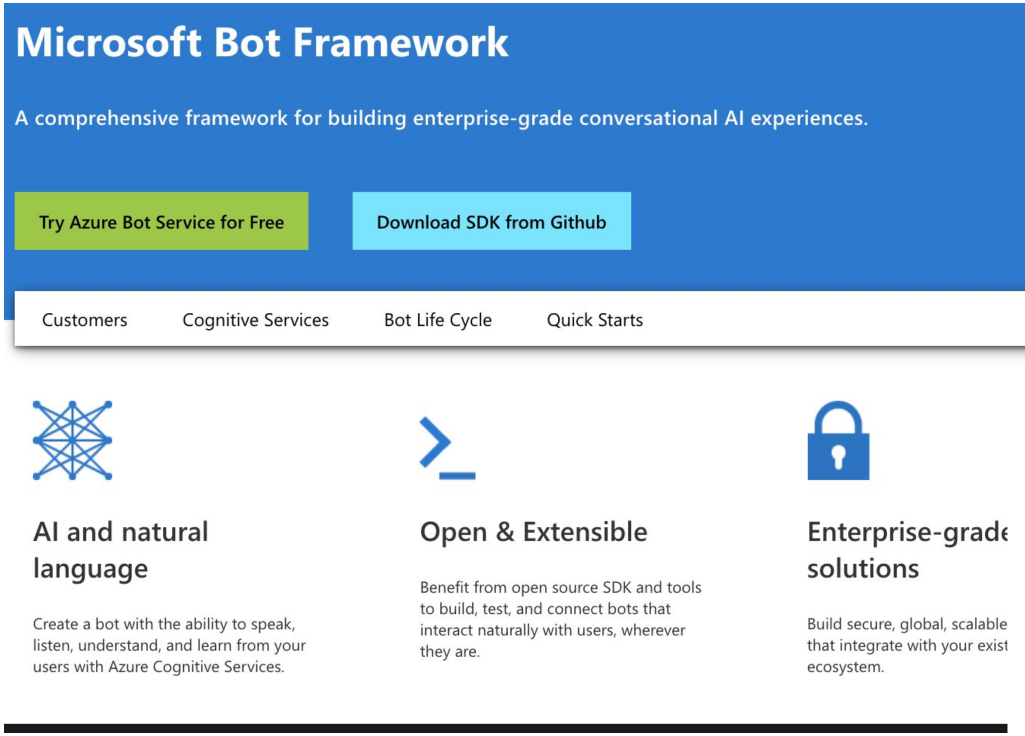
That action was simply to prevent Google Photos from ever labelling any image as a gorilla, chimpanzee, or monkey – even pictures of the primates themselves.

https://www.bbc.com/news/technology-33347866
https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people



Microsoft deletes 'teen girl' AI after it became a Hitler-loving sex robot within 24 hours

https://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/
https://www.bbc.com/news/technology-52978191



Microsoft Bot Framework

A comprehensive framework for building enterprise-grade conversational AI experiences.

AI and natural language

Create a bot with the ability to speak, listen, understand, and learn from your users with Azure Cognitive Services.

Open & Extensible

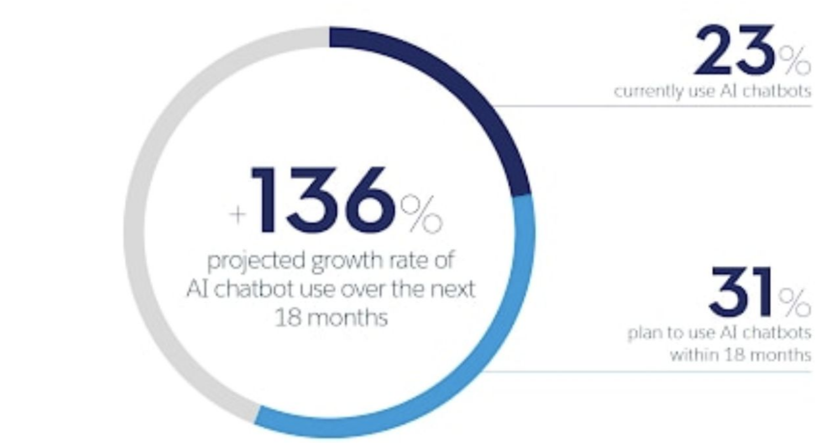
Benefit from open source SDK and tools to build, test, and connect bots that interact naturally with users, wherever they are.

Enterprise-grade solutions

Build secure, global, scalable that integrate with your exist ecosystem.

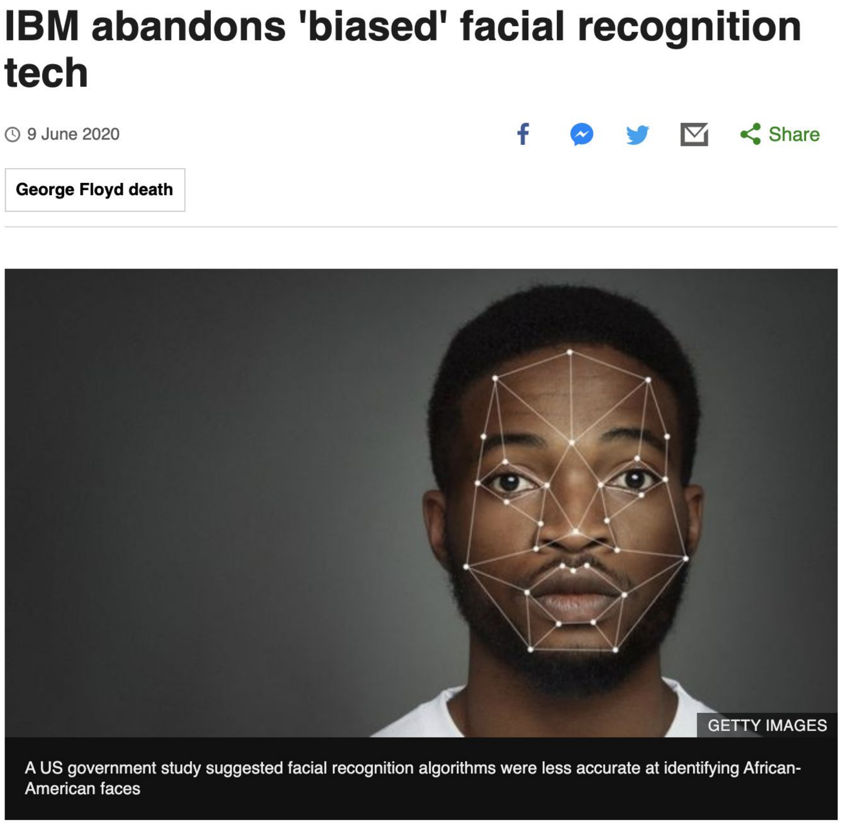
https://dev.botframework.com/

Use or Planned Use of AI Chatbots Among Service Organizations



"State of Service," Salesforce Research, March 2019.

https://www.salesforce.com/blog/chatbot-statistics/



IBM abandons 'biased' facial recognition tech

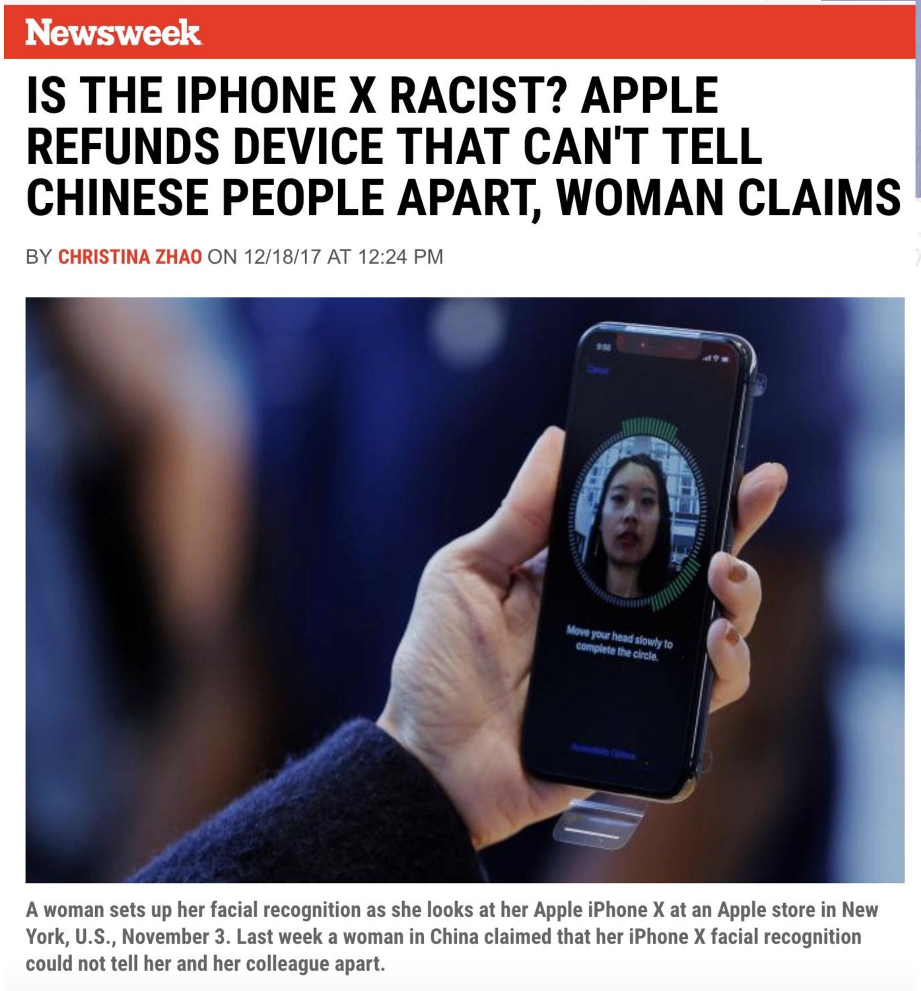
9 June 2020

Share

George Floyd death

GETTY IMAGES

A US government study suggested facial recognition algorithms were less accurate at identifying African-American faces

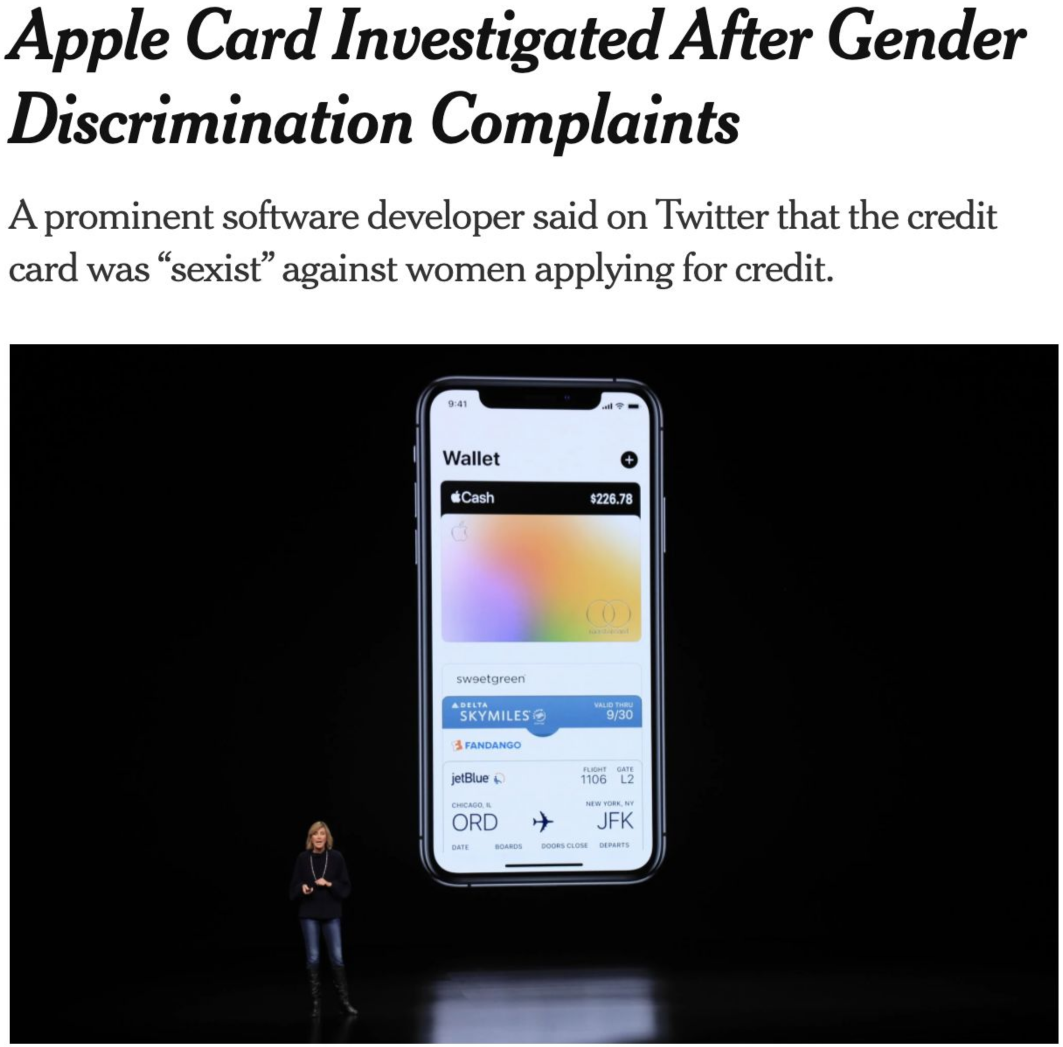


Newsweek

IS THE IPHONE X RACIST? APPLE REFUNDS DEVICE THAT CAN'T TELL CHINESE PEOPLE APART, WOMAN CLAIMS

BY CHRISTINA ZHAO ON 12/18/17 AT 12:24 PM

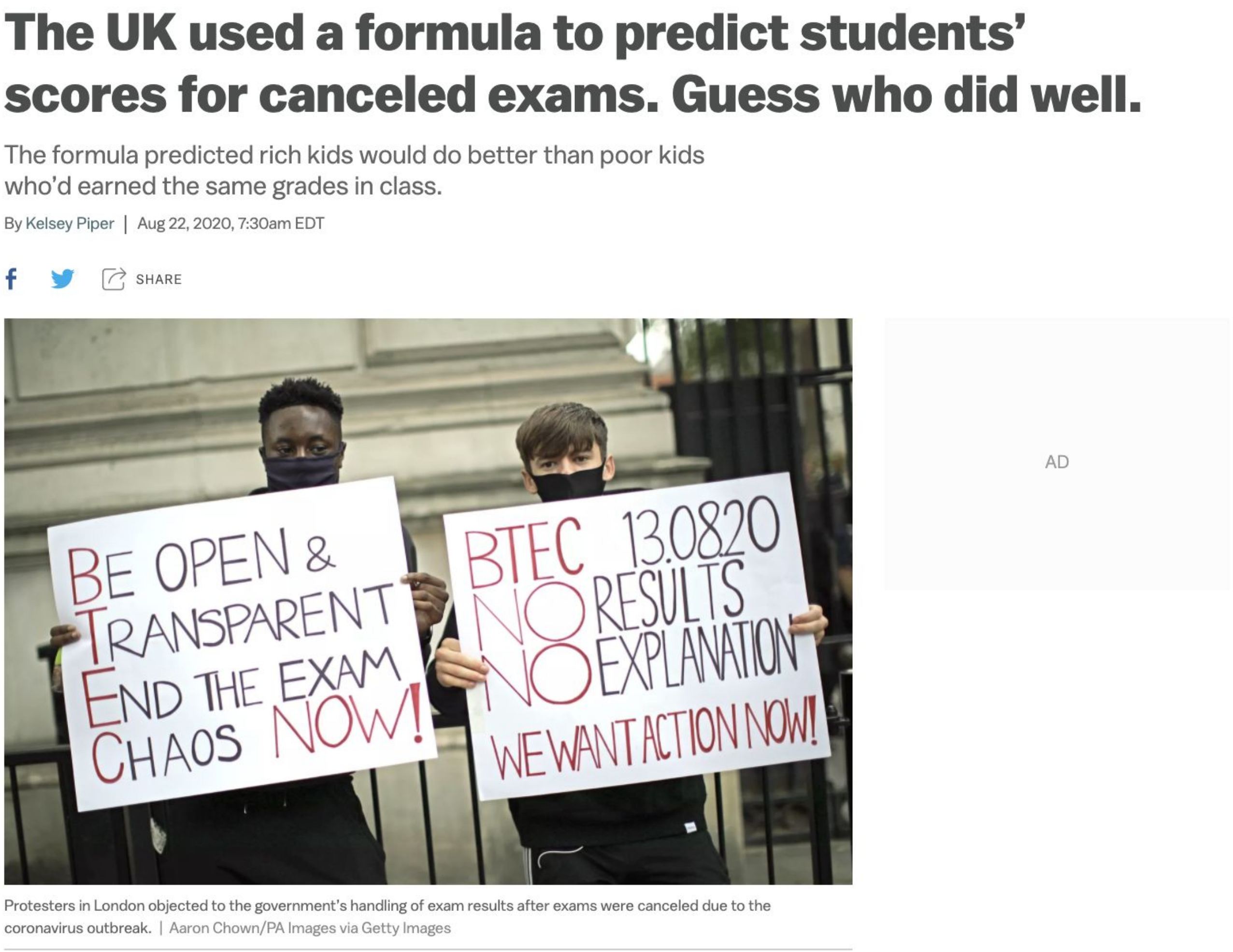
A woman sets up her facial recognition as she looks at her Apple iPhone X at an Apple store in New York, U.S., November 3. Last week a woman in China claimed that her iPhone X facial recognition could not tell her and her colleague apart.



Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was “sexist” against women applying for credit.

https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html



The UK used a formula to predict students' scores for canceled exams. Guess who did well.

The formula predicted rich kids would do better than poor kids who'd earned the same grades in class.

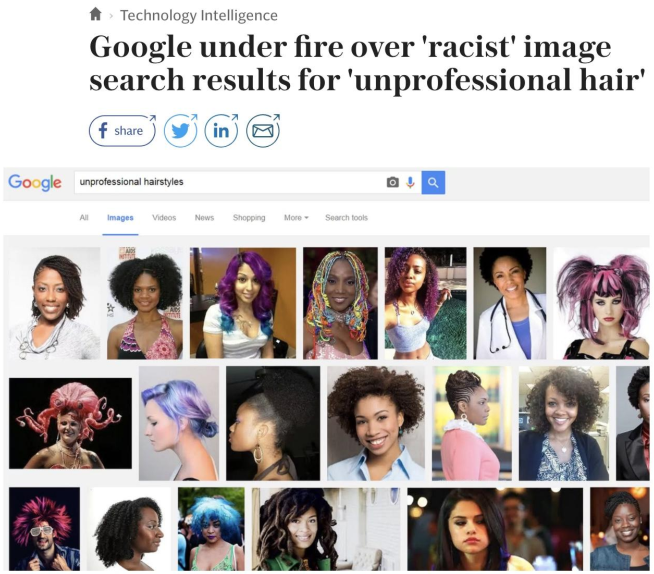
By Kelsey Piper | Aug 22, 2020, 7:30am EDT

SHARE
BE OPEN & TRANSPARENT END THE EXAM CHAOS NOW!
BTEC 13.08.20 NO RESULTS NO EXPLANATION WE WANT ACTION NOW!
AD

Protesters in London objected to the government's handling of exam results after exams were canceled due to the coronavirus outbreak. | Aaron Chown/PA Images via Getty Images

https://www.vox.com/future-perfect/2020/8/22/21374872/uk-united-kingdom-formula-predict-student-test-scores-exams

Bias in Information Access?



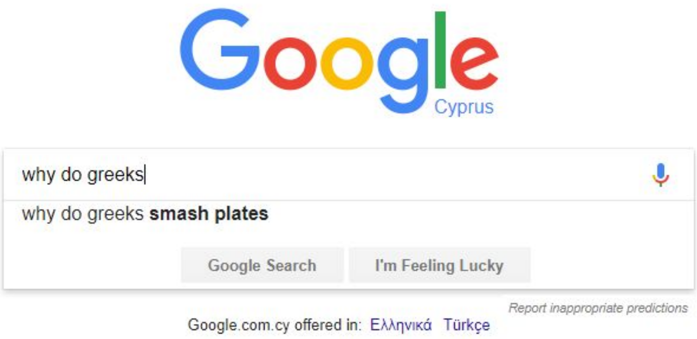
Technology Intelligence

Google under fire over 'racist' image search results for 'unprofessional hair'

unprofessional hairstyles

Google Image search results for 'unprofessional hair'

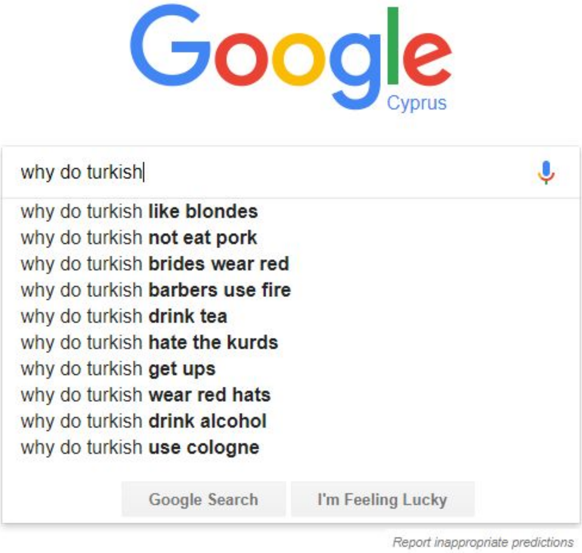
Bias in Information Access?



why do greeks

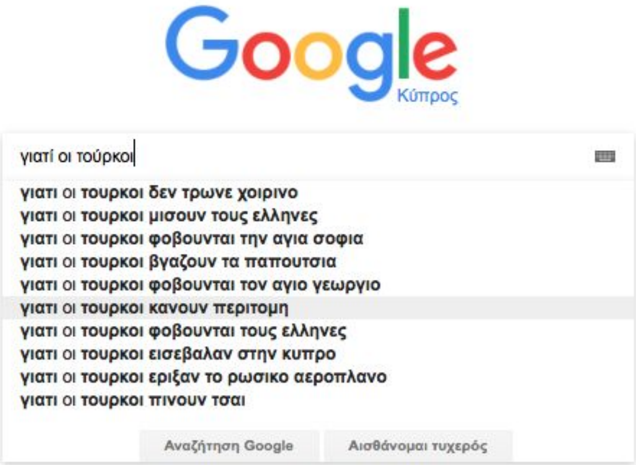
why do greeks **smash plates**

Google.com.cy offered in: Ελληνικά Türkçe



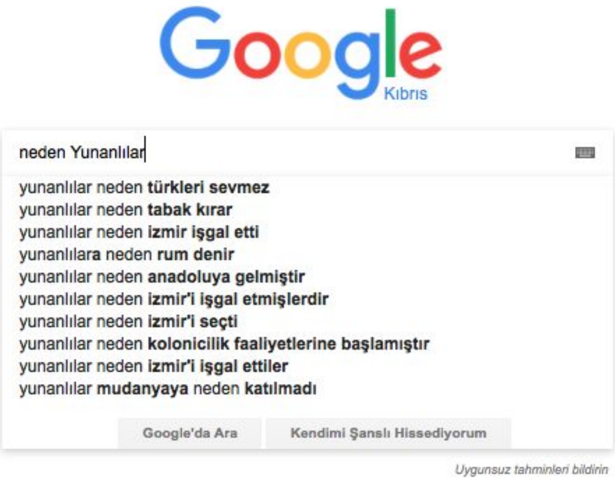
why do turkish

why do turkish **like blondes**
why do turkish **not eat pork**
why do turkish **brides wear red**
why do turkish **barbers use fire**
why do turkish **drink tea**
why do turkish **hate the kurds**
why do turkish **get ups**
why do turkish **wear red hats**
why do turkish **drink alcohol**
why do turkish **use cologne**



γιατί οι τούρκοι

γιατι οι **τουρκοι δεν τρωνε χοιρινο**
γιατι οι **τουρκοι μισουν τους ελληνες**
γιατι οι **τουρκοι φοβουνται την αγια σοφια**
γιατι οι **τουρκοι βγαζουν τα παπουτσια**
γιατι οι **τουρκοι φοβουνται τον αγιο γεωργιο**
γιατι οι **τουρκοι κανουν περιτομη**
γιατι οι **τουρκοι φοβουνται τους ελληνες**
γιατι οι **τουρκοι εισεβαλαν στην κυπρο**
γιατι οι **τουρκοι εριξαν το ρωσικο αεροπλανο**
γιατι οι **τουρκοι πινουν τσαι**



neden Yunanlılar

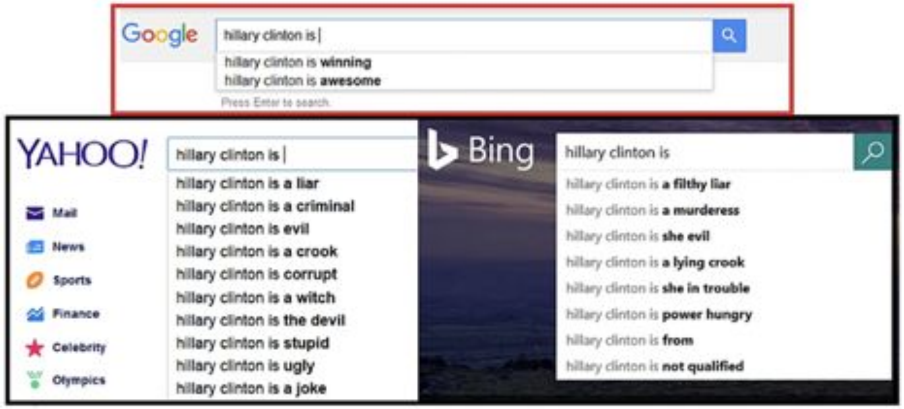
yunanlılar neden **türkleri sevmez**
yunanlılar neden **tabak kırar**
yunanlılar neden **izmir işgal etti**
yunanlıla**ra** neden **rum denir**
yunanlılar neden **anadoluya gelmiştir**
yunanlılar neden **izmir'i işgal etmişlerdir**
yunanlılar neden **izmir'i seçti**
yunanlılar neden **kolonicilik faaliyetlerine başlamıştır**
yunanlılar neden **izmir'i işgal ettiler**
yunanlılar **mudanyaya** neden **katılmadı**

All Systems have a Slant

Bias in information system is not a new problem!

1. Results are slanted in *unfair discrimination* against particular persons or groups
2. That discrimination is *systematic*

[Friedman & Nissenbaum, 1996]



hillary clinton is |

hillary clinton is **winning**
hillary clinton is **awesome**

Press Enter to search.

hillary clinton is |

hillary clinton is a liar
hillary clinton is a criminal
hillary clinton is evil
hillary clinton is a crook
hillary clinton is corrupt
hillary clinton is a witch
hillary clinton is the devil
hillary clinton is stupid
hillary clinton is ugly
hillary clinton is a joke

hillary clinton is

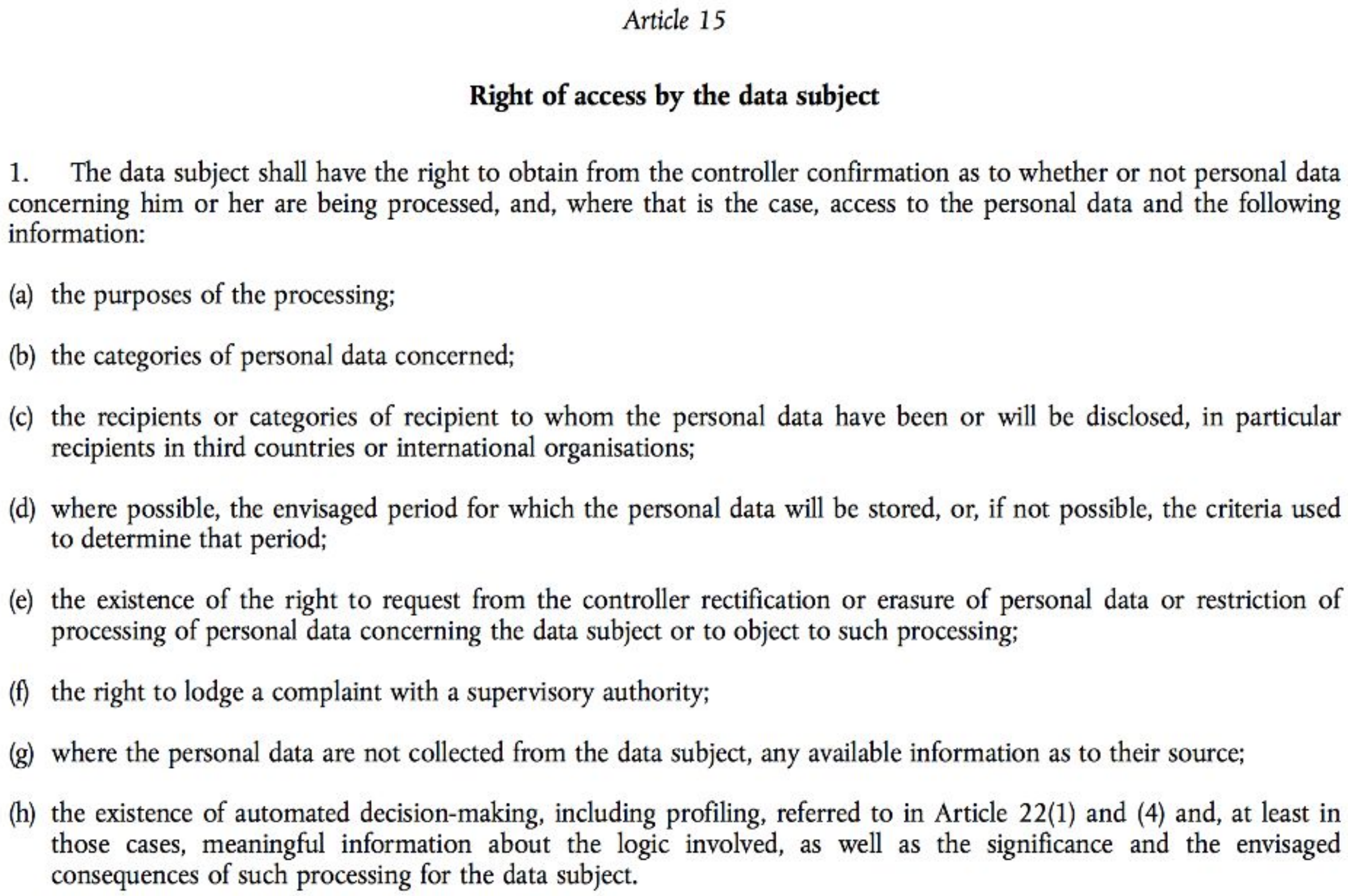
hillary clinton is **a filthy liar**
hillary clinton is **a murderess**
hillary clinton is **she evil**
hillary clinton is **a lying crook**
hillary clinton is **she in trouble**
hillary clinton is **power hungry**
hillary clinton is **from**
hillary clinton is **not qualified**

Response: Government / Regulators

EU: General Data Protection Regulation

- Is there a “right to an explanation”?
  - The right not to be subject to automated decision-making and safeguards enacted thereof (Article 22, Recital 71)
  - Notification duties of data controllers (Articles 13-14, Recitals 60-62)
  - The right to access (Article 15, Recital 63)

EU: GDPR



Article 15

Right of access by the data subject

1. The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:

(a) the purposes of the processing;

(b) the categories of personal data concerned;

(c) the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;

(d) where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;

(e) the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;

(f) the right to lodge a complaint with a supervisory authority;

(g) where the personal data are not collected from the data subject, any available information as to their source;

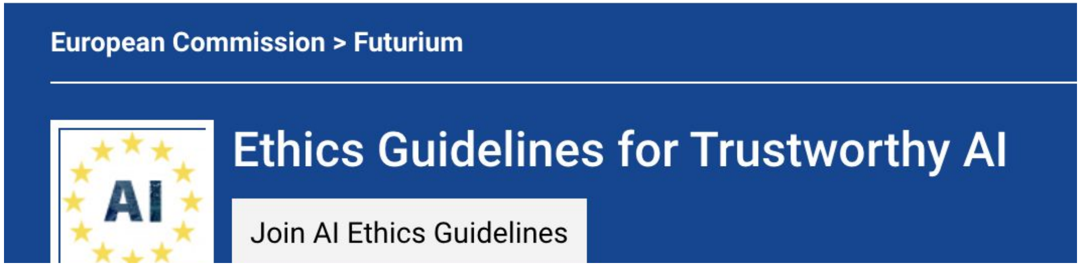
(h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

EU: GDPR

Just a few challenges…

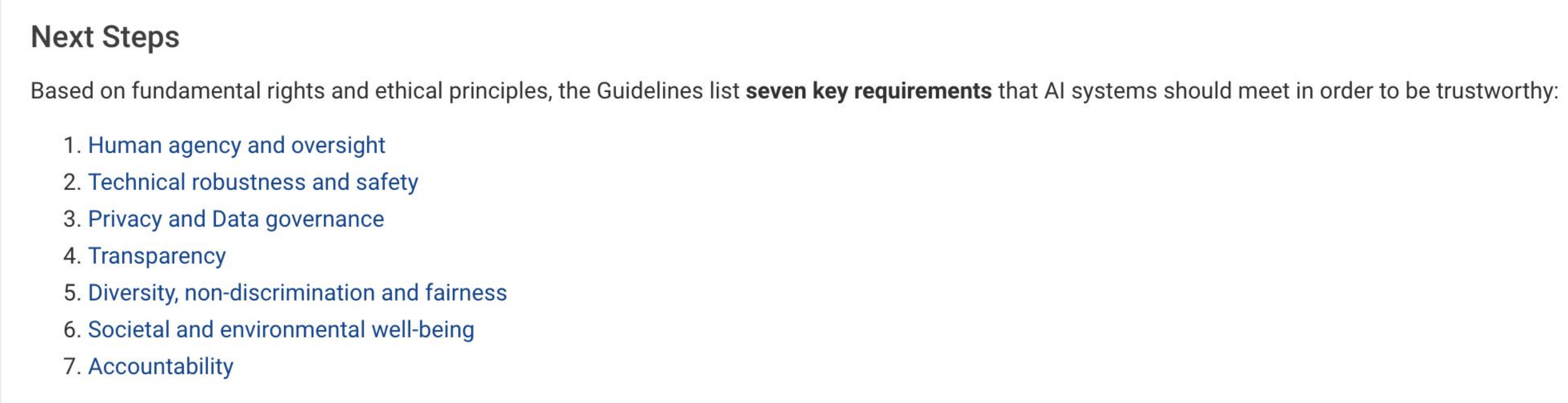
- Vague language
  - “meaningful information/explanation”
  - “logic involved”
  - “significance”
  - “envisaged consequences”
- What kinds of “meaningful explanations”?
  - Global vs. local explanations
  - Explanation for whom?
  - Issues of algorithmic and digital literacy

EC: Trustworthy AI



European Commission > Futurium

Ethics Guidelines for Trustworthy AI

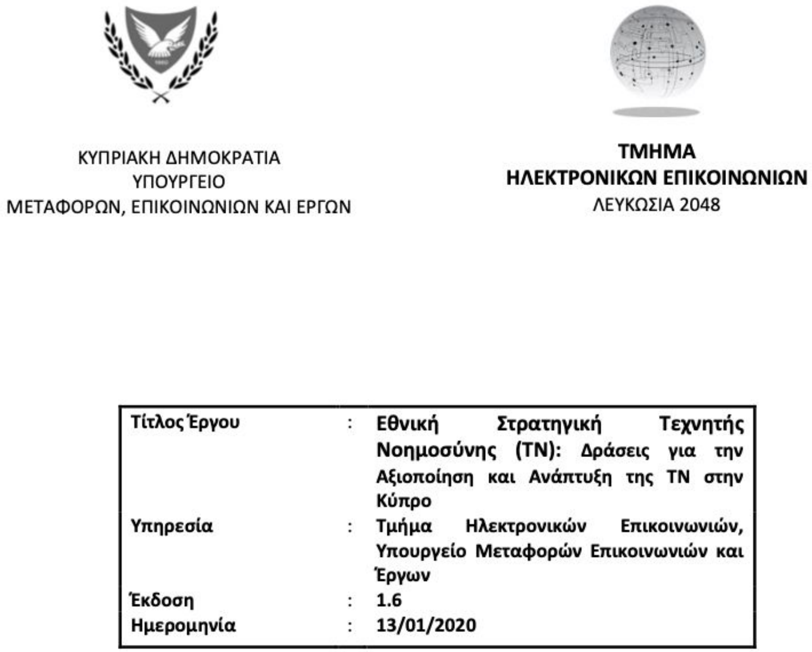


Next Steps

Based on fundamental rights and ethical principles, the Guidelines list **seven key requirements** that AI systems should meet in order to be trustworthy:

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and Data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

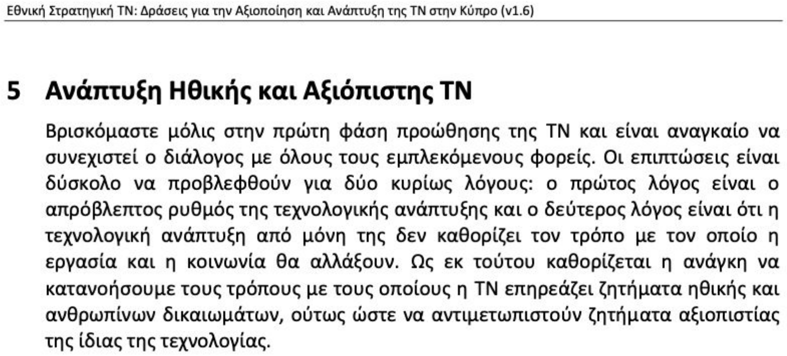
National AI Strategies



ΚΥΠΡΙΑΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ
ΜΕΤΑΦΟΡΩΝ, ΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΕΡΓΩΝ

ΤΜΗΜΑ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΕΠΙΚΟΙΝΩΝΙΩΝ
ΛΕΥΚΩΣΙΑ 2048

Τίτλος Έργου	:	**Εθνική Στρατηγική Τεχνητής Νοημοσύνης (ΤΝ): Δράσεις για την Αξιοποίηση και Ανάπτυξη της ΤΝ στην Κύπρο**
Υπηρεσία	:	**Τμήμα Ηλεκτρονικών Επικοινωνιών, Υπουργείο Μεταφορών Επικοινωνιών και Έργων**
Έκδοση	:	**1.6**
Ημερομηνία	:	**13/01/2020**

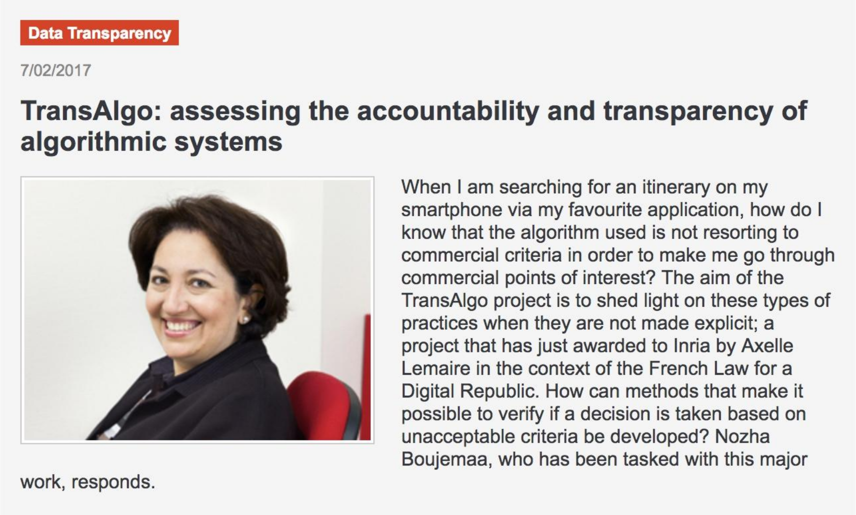


Εθνική Στρατηγική ΤΝ: Δράσεις για την Αξιοποίηση και Ανάπτυξη της ΤΝ στην Κύπρο (v1.6)

5 Ανάπτυξη Ηθικής και Αξιόπιστης ΤΝ

Βρισκόμαστε μόλις στην πρώτη φάση προώθησης της ΤΝ και είναι αναγκαίο να συνεχιστεί ο διάλογος με όλους τους εμπλεκόμενους φορείς. Οι επιπτώσεις είναι δύσκολο να προβλεφθούν για δύο κυρίως λόγους: ο πρώτος λόγος είναι ο απρόβλεπτος ρυθμός της τεχνολογικής ανάπτυξης και ο δεύτερος λόγος είναι ότι η τεχνολογική ανάπτυξη από μόνη της δεν καθορίζει τον τρόπο με τον οποίο η εργασία και η κοινωνία θα αλλάξουν. Ως εκ τούτου καθορίζεται η ανάγκη να κατανοήσουμε τους τρόπους με τους οποίους η ΤΝ επηρεάζει ζητήματα ηθικής και ανθρωπίνων δικαιωμάτων, ούτως ώστε να αντιμετωπιστούν ζητήματα αξιοπιστίας της ίδιας της τεχνολογίας.

National Level

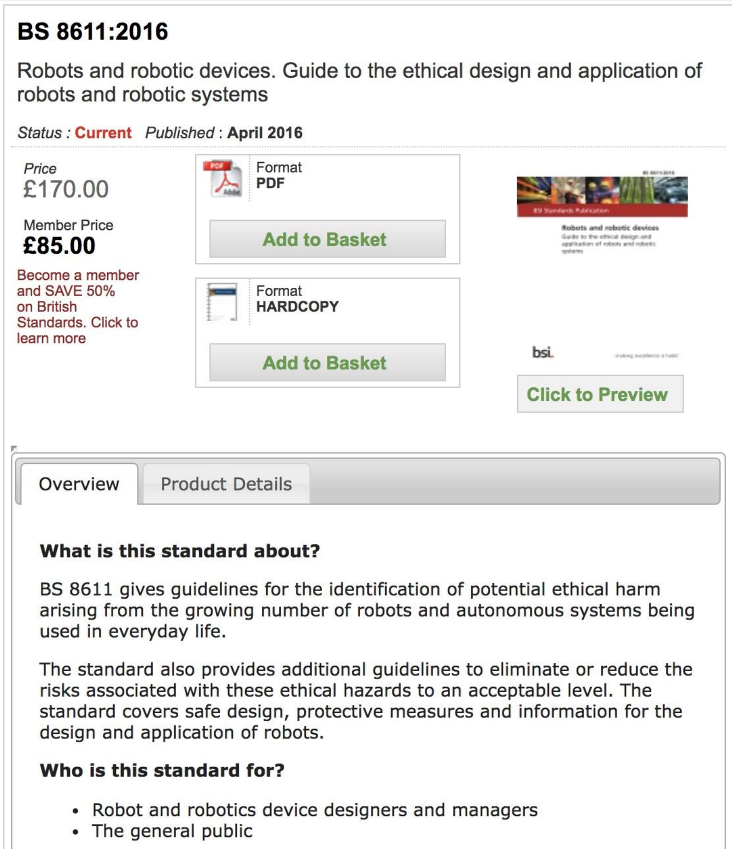


Data Transparency

7/02/2017

TransAlgo: assessing the accountability and transparency of algorithmic systems

When I am searching for an itinerary on my smartphone via my favourite application, how do I know that the algorithm used is not resorting to commercial criteria in order to make me go through commercial points of interest? The aim of the TransAlgo project is to shed light on these types of practices when they are not made explicit; a project that has just awarded to Inria by Axelle Lemaire in the context of the French Law for a Digital Republic. How can methods that make it possible to verify if a decision is taken based on unacceptable criteria be developed? Nozha Boujemaa, who has been tasked with this major work, responds.



BS 8611:2016

Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems

Status : **Current** *Published :* **April 2016**

Price
£170.00

Member Price
£85.00

Become a member and SAVE 50% on British Standards. Click to learn more

Format
PDF

Add to Basket

Format
HARDCOPY

Add to Basket

Click to Preview

Overview | Product Details

What is this standard about?

BS 8611 gives guidelines for the identification of potential ethical harm arising from the growing number of robots and autonomous systems being used in everyday life.

The standard also provides additional guidelines to eliminate or reduce the risks associated with these ethical hazards to an acceptable level. The standard covers safe design, protective measures and information for the design and application of robots.

Who is this standard for?

- Robot and robotics device designers and managers
- The general public

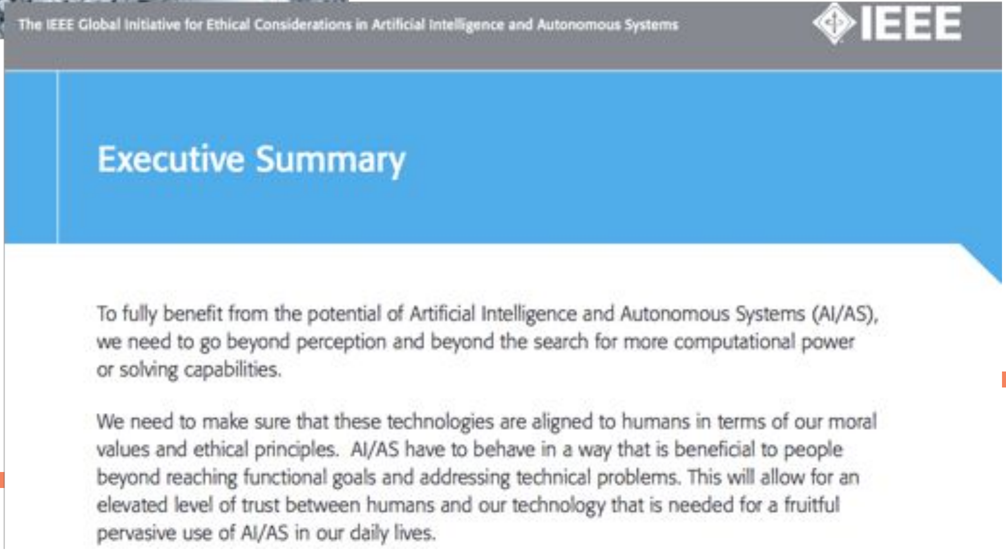
Response: Industry & Professions

OVERVIEW

IEEE
Advancing Technology for Humanity

ETHICALLY ALIGNED DESIGN

A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems



The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

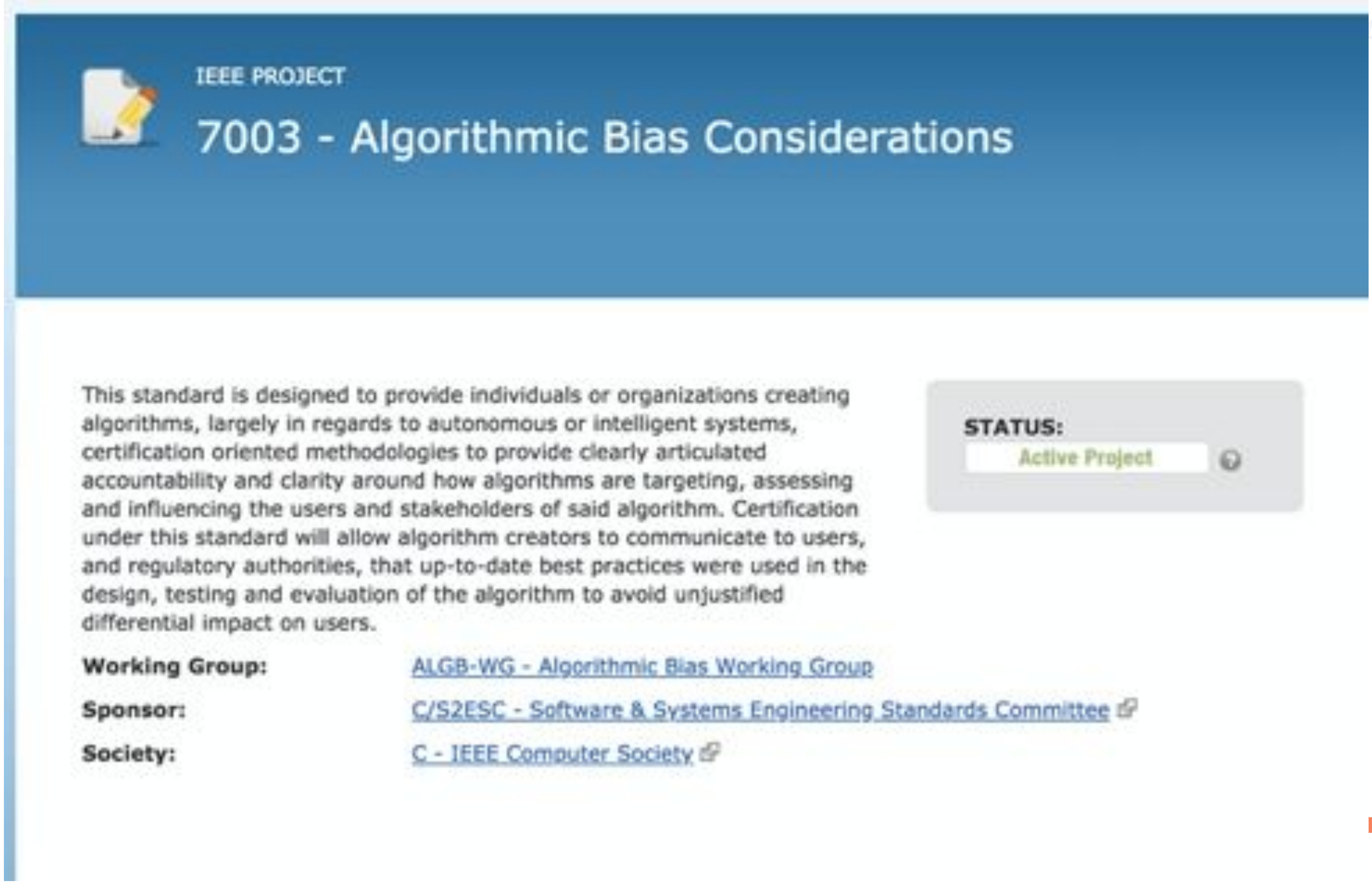
IEEE

Executive Summary

To fully benefit from the potential of Artificial Intelligence and Autonomous Systems (AI/AS), we need to go beyond perception and beyond the search for more computational power or solving capabilities.

We need to make sure that these technologies are aligned to humans in terms of our moral values and ethical principles. AI/AS have to behave in a way that is beneficial to people beyond reaching functional goals and addressing technical problems. This will allow for an elevated level of trust between humans and our technology that is needed for a fruitful pervasive use of AI/AS in our daily lives.

IEEE 7003



IEEE PROJECT

7003 - Algorithmic Bias Considerations

This standard is designed to provide individuals or organizations creating algorithms, largely in regards to autonomous or intelligent systems, certification oriented methodologies to provide clearly articulated accountability and clarity around how algorithms are targeting, assessing and influencing the users and stakeholders of said algorithm. Certification under this standard will allow algorithm creators to communicate to users, and regulatory authorities, that up-to-date best practices were used in the design, testing and evaluation of the algorithm to avoid unjustified differential impact on users.

STATUS:
Active Project

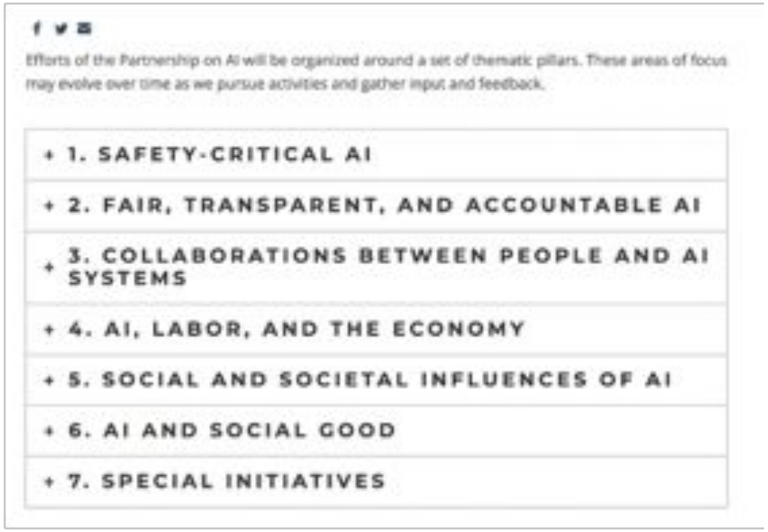
Working Group:	ALGB-WG - Algorithmic Bias Working Group
Sponsor:	C/S2ESC - Software & Systems Engineering Standards Committee
Society:	C - IEEE Computer Society

Industry Partnerships

Partnership on AI

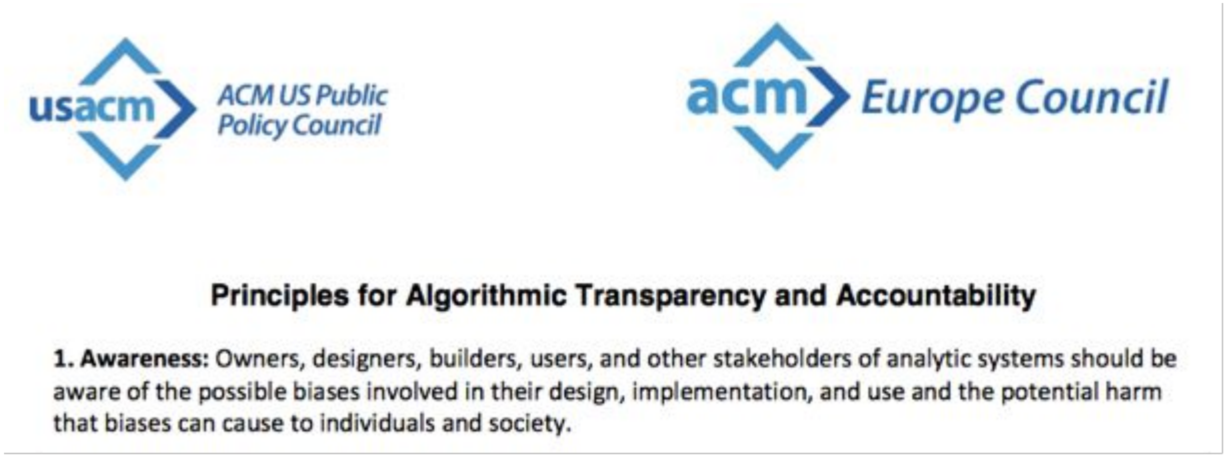
to benefit people and society

Established to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society.



Efforts of the Partnership on AI will be organized around a set of thematic pillars. These areas of focus may evolve over time as we pursue activities and gather input and feedback.

- + 1. SAFETY-CRITICAL AI
- + 2. FAIR, TRANSPARENT, AND ACCOUNTABLE AI
- + 3. COLLABORATIONS BETWEEN PEOPLE AND AI SYSTEMS
- + 4. AI, LABOR, AND THE ECONOMY
- + 5. SOCIAL AND SOCIETAL INFLUENCES OF AI
- + 6. AI AND SOCIAL GOOD
- + 7. SPECIAL INITIATIVES

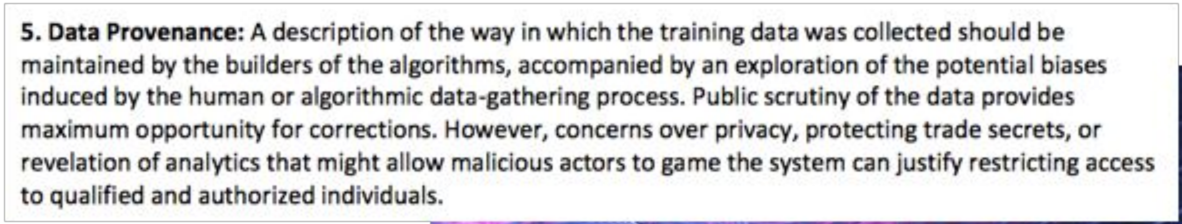


usacm ACM US Public Policy Council

acm Europe Council

Principles for Algorithmic Transparency and Accountability

1. Awareness: Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.



5. Data Provenance: A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

Break (15 minutes)

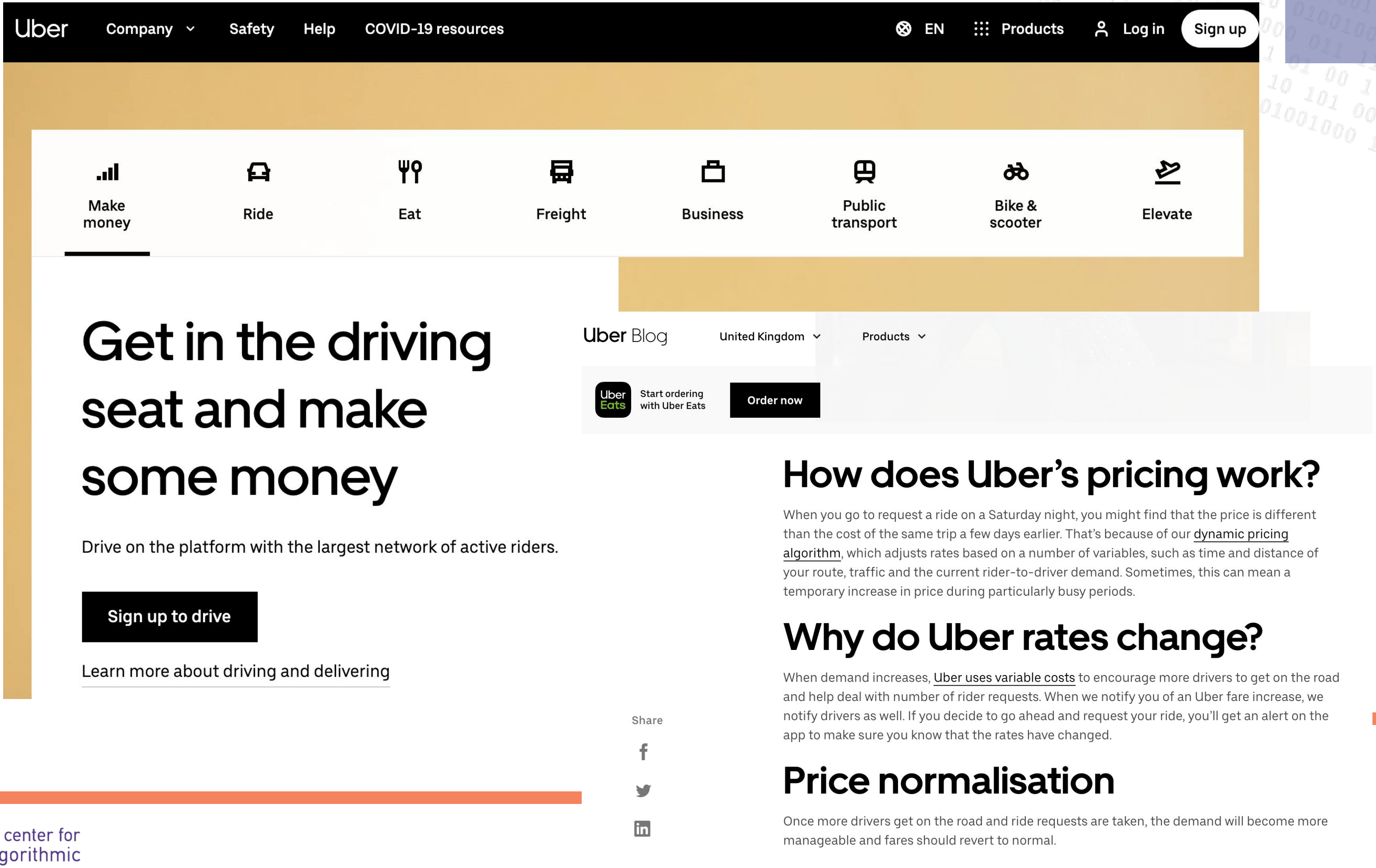
FATE as a Scientific Field

Background: FATE Research

- Some illustrative examples

  - Uber
    Dynamic pricing algorithms
  - Fiverr & TaskRabbit freelance marketplaces
    Recommendation systems
  - Search engines
    Information access (ranking, personalization)
  - Image tagging APIs
    Computer vision

UBER



Get in the driving seat and make some money

Drive on the platform with the largest network of active riders.

Learn more about driving and delivering

How does Uber's pricing work?

When you go to request a ride on a Saturday night, you might find that the price is different than the cost of the same trip a few days earlier. That's because of our dynamic pricing algorithm, which adjusts rates based on a number of variables, such as time and distance of your route, traffic and the current rider-to-driver demand. Sometimes, this can mean a temporary increase in price during particularly busy periods.

Why do Uber rates change?

When demand increases, Uber uses variable costs to encourage more drivers to get on the road and help deal with number of rider requests. When we notify you of an Uber fare increase, we notify drivers as well. If you decide to go ahead and request your ride, you'll get an alert on the app to make sure you know that the rates have changed.

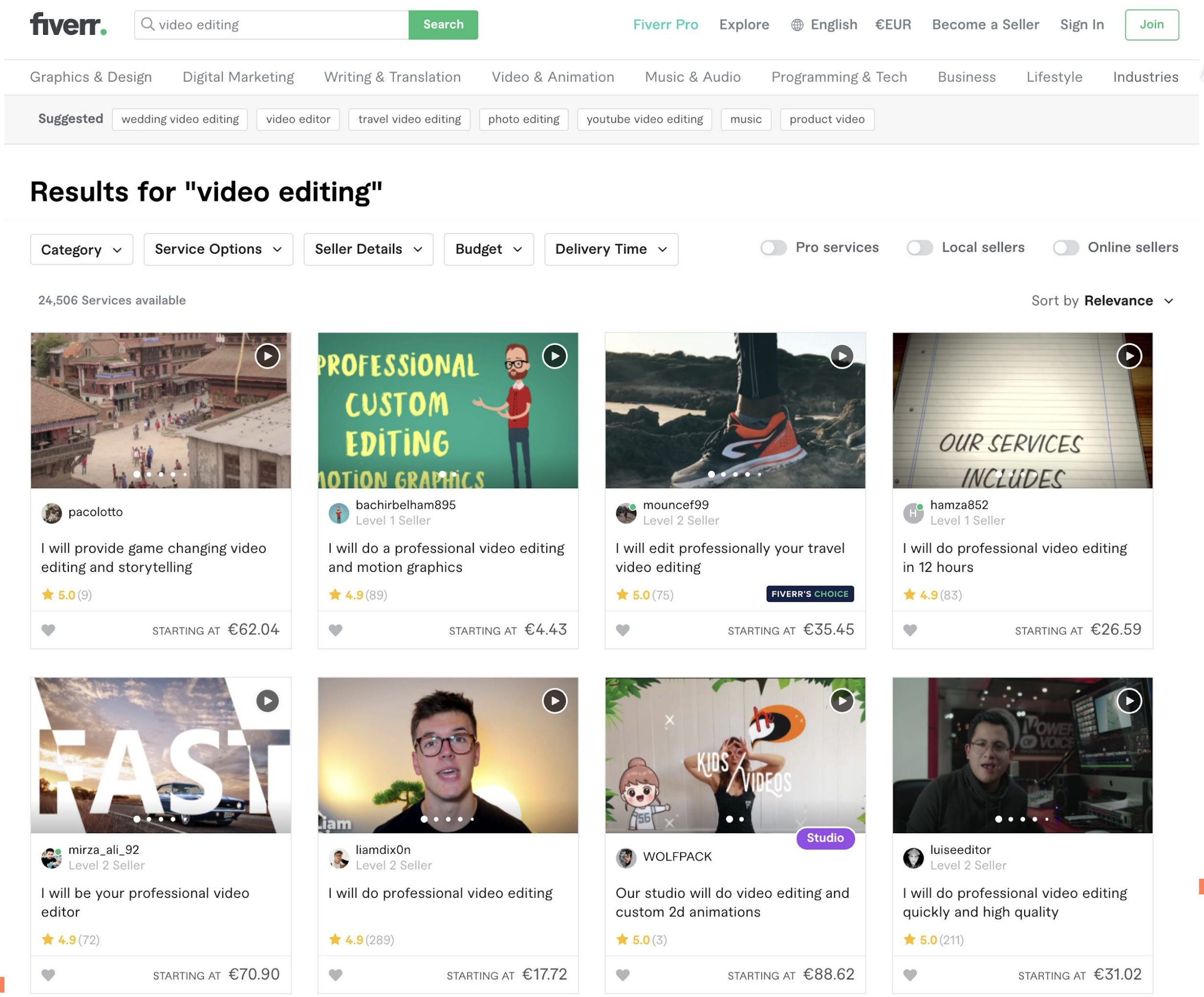
Price normalisation

Once more drivers get on the road and ride requests are taken, the demand will become more manageable and fares should revert to normal.

Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber’s drivers. *International Journal of Communication*, *10*, 27.

- “Uber’s claims regarding its labor model, which center on *freedom, flexibility, and entrepreneurship*, are complicated and contradicted by the experience of its drivers.”
- “Power and information asymmetries emerge via Uber’s software-based platform through algorithmic labor logistics *shaping driver behavior*, electronic surveillance, and policies for performance targets. “
- “Through the Uber app’s design and deployment, the company produces the equivalent effects of what most reasonable observers would define as a *managed labor force*.”

Fiverr



Results for "video editing"

24,506 Services available

Sort by **Relevance**

pacolotto
I will provide game changing video editing and storytelling
★ 5.0 (9)
STARTING AT €62.04

bachirbelham895
Level 1 Seller
I will do a professional video editing and motion graphics
★ 4.9 (89)
STARTING AT €4.43

mouncef99
Level 2 Seller
I will edit professionally your travel video editing
★ 5.0 (75)
FIVERR'S CHOICE
STARTING AT €35.45

hamza852
Level 1 Seller
I will do professional video editing in 12 hours
★ 4.9 (83)
STARTING AT €26.59

mirza_ali_92
Level 2 Seller
I will be your professional video editor
★ 4.9 (72)
STARTING AT €70.90

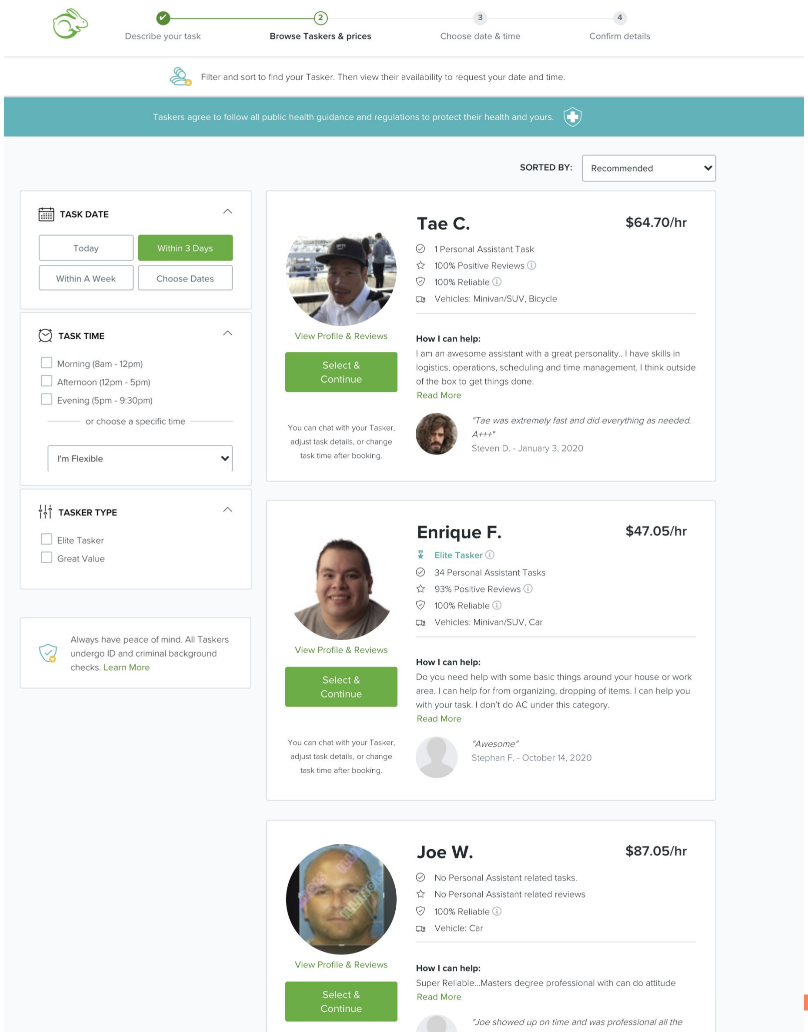
liamdix0n
Level 2 Seller
I will do professional video editing
★ 4.9 (289)
STARTING AT €17.72

WOLFPACK
Our studio will do video editing and custom 2d animations
★ 5.0 (3)
STARTING AT €88.62

luiseeditor
Level 2 Seller
I will do professional video editing quickly and high quality
★ 5.0 (211)
STARTING AT €31.02

CY. center for algorithmic transparency

Task Rabbit



✓ Describe your task — 2 Browse Taskers & prices — 3 Choose date & time — 4 Confirm details

Filter and sort to find your Tasker. Then view their availability to request your date and time.

Taskers agree to follow all public health guidance and regulations to protect their health and yours.

SORTED BY: Recommended

TASK DATE

Today | **Within 3 Days**

Within A Week | Choose Dates

TASK TIME

□ Morning (8am - 12pm)
□ Afternoon (12pm - 5pm)
□ Evening (5pm - 9:30pm)

or choose a specific time

I'm Flexible

TASKER TYPE

□ Elite Tasker
□ Great Value

Always have peace of mind. All Taskers undergo ID and criminal background checks. Learn More

Tae C. $64.70/hr

- 1 Personal Assistant Task
- 100% Positive Reviews
- 100% Reliable
- Vehicles: Minivan/SUV, Bicycle

View Profile & Reviews

Select & Continue

You can chat with your Tasker, adjust task details, or change task time after booking.

How I can help:
I am an awesome assistant with a great personality.. I have skills in logistics, operations, scheduling and time management. I think outside of the box to get things done.
Read More

"Tae was extremely fast and did everything as needed. A+++"
Steven D. - January 3, 2020

Enrique F. $47.05/hr

- Elite Tasker
- 34 Personal Assistant Tasks
- 93% Positive Reviews
- 100% Reliable
- Vehicles: Minivan/SUV, Car

View Profile & Reviews

Select & Continue

You can chat with your Tasker, adjust task details, or change task time after booking.

How I can help:
Do you need help with some basic things around your house or work area. I can help for from organizing, dropping of items. I can help you with your task. I don't do AC under this category.
Read More

"Awesome"
Stephan F. - October 14, 2020

Joe W. $87.05/hr

- No Personal Assistant related tasks.
- No Personal Assistant related reviews
- 100% Reliable
- Vehicle: Car

View Profile & Reviews

Select & Continue

How I can help:
Super Reliable...Masters degree professional with can do attitude
Read More

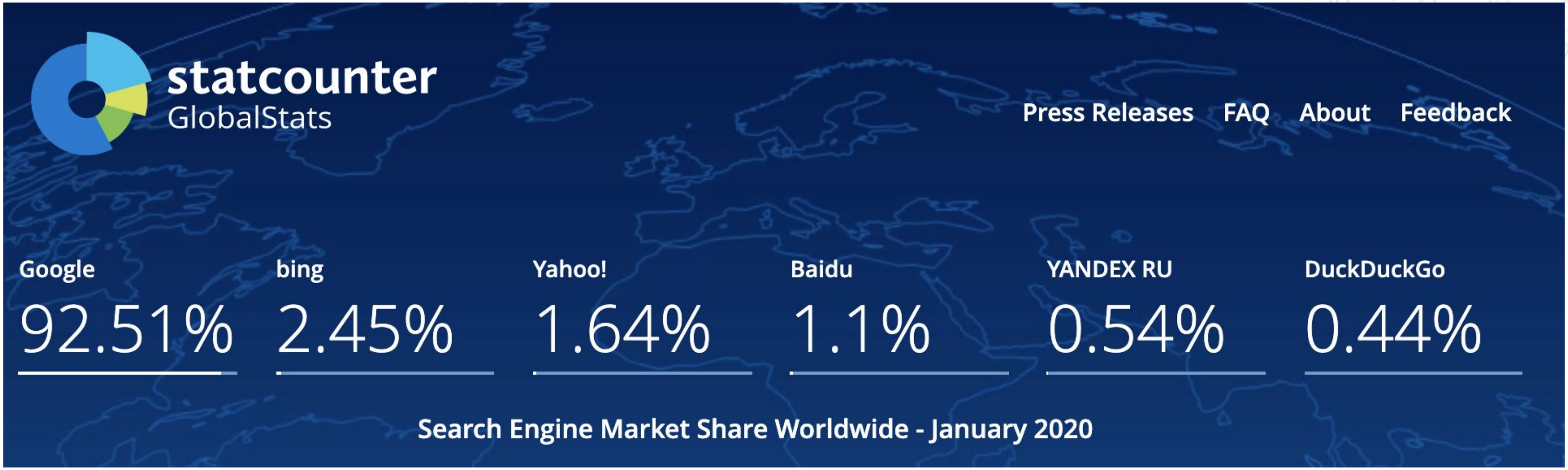
"Joe showed up on time and was professional all the

Hannák, A., Wagner, C., Garcia, D., Mislove, A., Strohmaier, M., & Wilson, C. (2017, February). Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 1914-1933).

Audit of worker rankings & reviews

- "Workers perceived to be women, especially White women, receive 10% fewer reviews than workers perceived to be men with equivalent work experience."
- "Workers perceived to be Black, especially men, receive significantly lower feedback scores (i.e., ratings) than other workers with similar attributes."

Search Engine Bias(?)



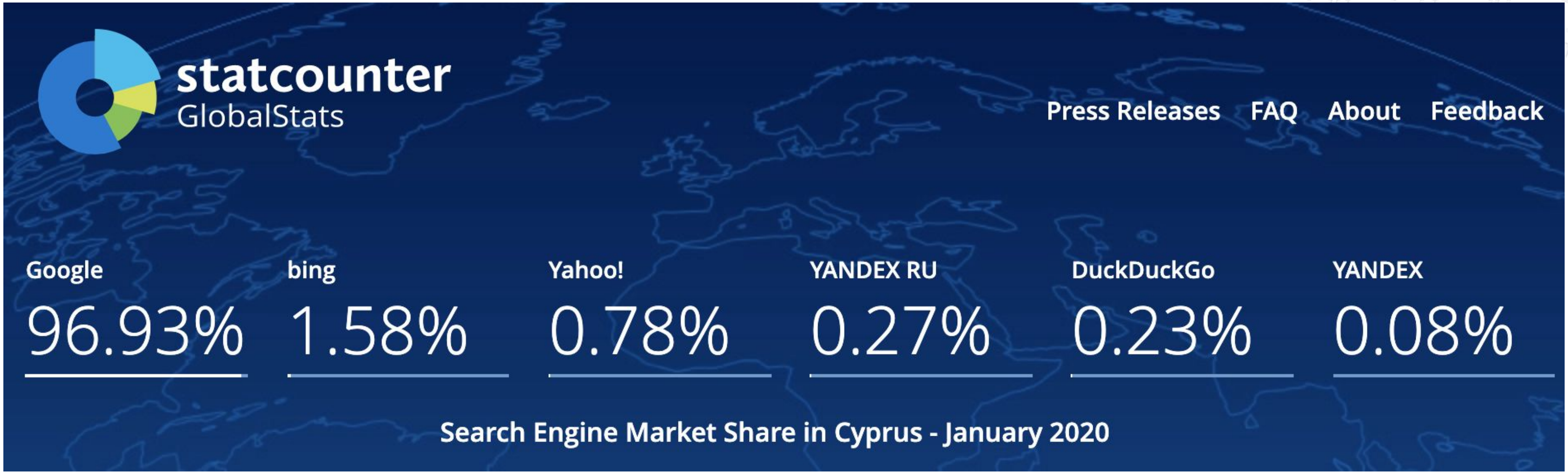
statcounter GlobalStats

Press Releases FAQ About Feedback

Google	bing	Yahoo!	Baidu	YANDEX RU	DuckDuckGo
92.51%	2.45%	1.64%	1.1%	0.54%	0.44%

Search Engine Market Share Worldwide - January 2020

Search Engine Bias(?)



Google	bing	Yahoo!	YANDEX RU	DuckDuckGo	YANDEX
96.93%	1.58%	0.78%	0.27%	0.23%	0.08%

Search Engine Market Share in Cyprus - January 2020

Mowshowitz, A., & Kawaguchi, A. (2005). Measuring search engine bias. *Information Processing & Management*, *41*(5), 1193-1205.

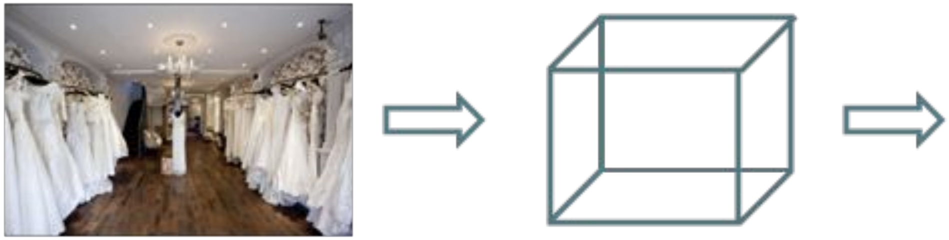
- Methodology for quantifying “bias” in search engine results, as a relative measure
- “The bias measure is designed to capture the degree to which the distribution of URLs, retrieved by a search engine in response to a query deviates from an idea of fair distribution for that query.”
- Experiments with 16 (!) search engines
- Main conclusion: lots of variance between engines, and by subject / topic

Image Tagging Algorithms

Input image

ITAs

Description as “tags”



- dressing_room
- clothing_store
- shop
- building
- cloakroom
- dress_rack
- closet
- clothing
- fabric
- grey_color

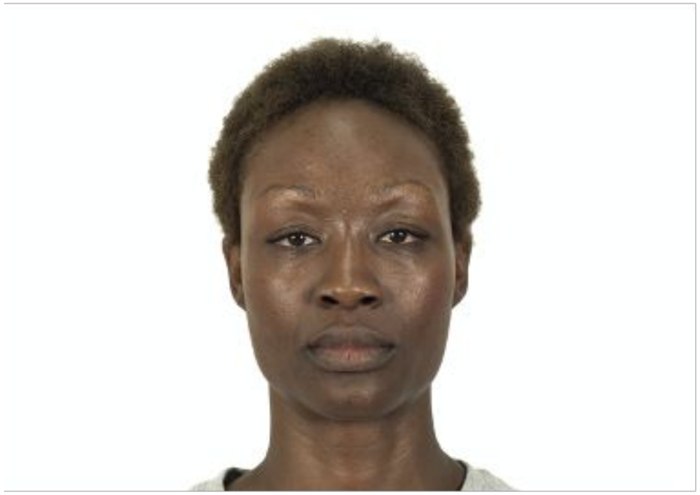
amazon web services Rekognition

clarifai

imagga

Microsoft Face API

IBM Watson



BF-231 from the Chicago Face Dataset, and tags output by the six image tagging APIs for this image

Amazon Rekognition	human, people, person, Afro, hairstyle, hair, face
Clarifai	people, one, portrait, man, wear, adult, side, pensive, profile, woman, face, isolated, child, facial, Afro, casual, fashion, athlete, adolescent
Google Cloud Vision	face, forehead, chin, eyebrow, cheek, nose, head, jaw, neck, human
Imagga Auto-tagger	Afro, man, face, portrait, male, handsome, head
Microsoft Vision	man, person, wearing, looking, necktie, standing, shirt, front, face, smiling, white, suit, posing, hair, holding, neck, young, glasses, black, head, hat, red
Watson Vision	person, woman, female, indian red color, coal black color

Are taggers “fair” towards the people in the images?

Group fairness: people from different protected classes (such as race and gender) should not experience significantly different treatment as compared to the majority or the population as a whole (Feldman et al. 2015)

Auditing the Black Boxes

Kyriakou, K., Barlas, P., Kleanthous, S., & Otterbacher, J. (2019, July). Fairness in proprietary image tagging algorithms: A cross-platform audit on people images. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 13, pp. 313-322).

Two approaches:

•**within-platform audits:** to discover how outputs may differ *for certain categories of inputs* in one system (e.g., Sweeney 2013)

•**cross-platform audits:** to discover how all outputs of *one system* may differ from *outputs of other systems*, for the same input
(e.g., Eslami et al. 2017)

Are Taggers Fair?
The Short Answer: No

- “Some [taggers] offer more interpretation on images, they may exhibit less fairness toward the depicted persons, by misuse of gender-related tags and/or making judgments on physical appearance.”
    - Asian females → more “attractiveness” tags
    - Black males → less interpretive tags

User Perception of Fairness

Barlas, P., Kleanthous, S., Kyriakou, K., & Otterbacher, J. (2019, June). What Makes an Image Tagger Fair?. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 95-103).

"Today, many automated tools are used to generate descriptions of images on the Web. However, some tools exhibit biases when processing images of people. Given an image and two descriptions of its content, decide which one is more fair."

"Imagine that auto-tagging is used to facilitate **searching profiles of people at a dating site**. Which of the above descriptions is **more fair**? Enter 0 if you cannot tell."

"Please **explain your answer regarding fairness**."

Description 1:

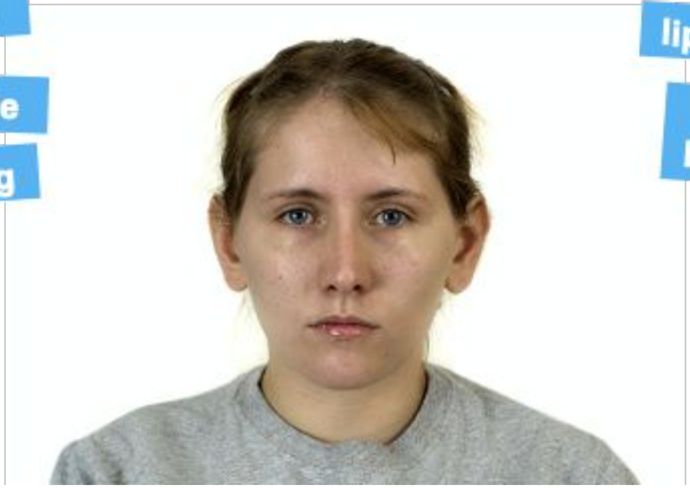
adolescent
contemporary
casual
cute
child
facial expression
eye
fine looking
fun
fashion
isolated
funny
man
one
looking
serious
people
portrait
wear
young

clarifai

Description 2:

blue eyes
woman
girl
caucasian
brown hair
front view
plain expression
short bangs
solo
sober
lip gloss
young
round face
white background
long hair

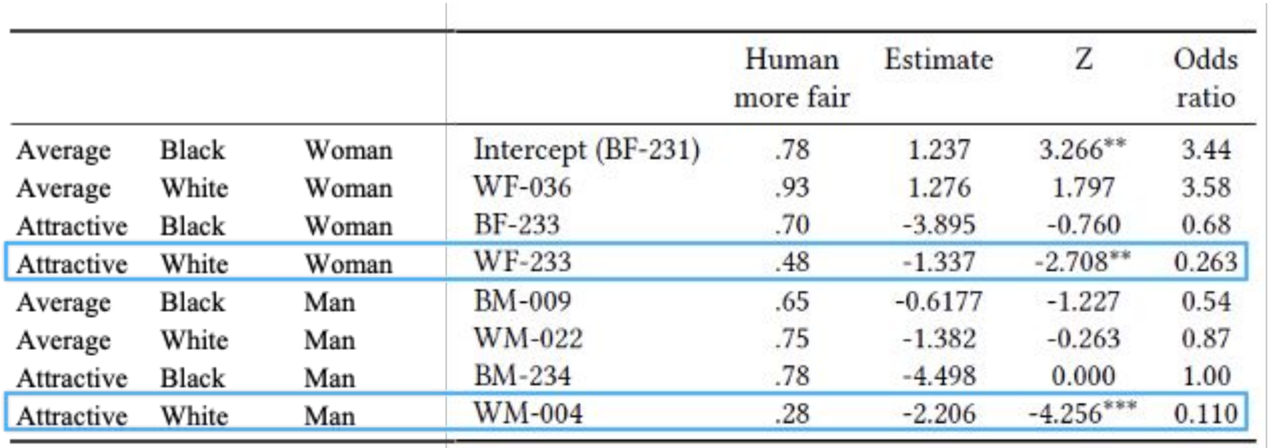
Crowdworkers



Experimental Set-up

Image	Gender	Race	“Attractiveness”	Participants (W/M)
BF-231	Woman	Black	Average	20/20
BF-233	Woman	Black	Attractive	20/20
WF-036	Woman	White	Average	20/20
WF-233	Woman	White	Attractive	20/20
BM-009	Man	Black	Average	20/20
BM-234	Man	Black	Attractive	20/20
WM-022	Man	White	Average	20/20
WM-004	Man	White	Attractive	20/20

Which is more “fair”: human or algorithm?

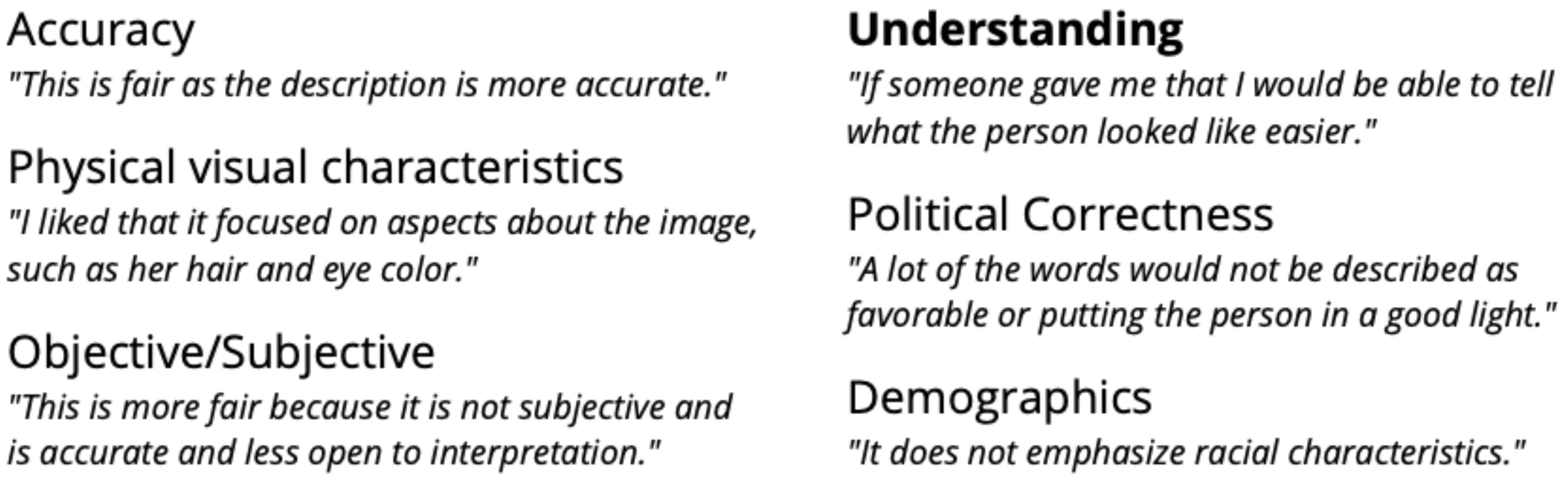


				Human more fair	Estimate	Z	Odds ratio
Average	Black	Woman	Intercept (BF-231)	.78	1.237	3.266**	3.44
Average	White	Woman	WF-036	.93	1.276	1.797	3.58
Attractive	Black	Woman	BF-233	.70	-3.895	-0.760	0.68
Attractive	White	Woman	WF-233	.48	-1.337	-2.708**	0.263
Average	Black	Man	BM-009	.65	-0.6177	-1.227	0.54
Average	White	Man	WM-022	.75	-1.382	-0.263	0.87
Attractive	Black	Man	BM-234	.78	-4.498	0.000	1.00
Attractive	White	Man	WM-004	.28	-2.206	-4.256***	0.110

Logit model to predict the event that human-generated tags are perceived as being more fair.

*** $p < .001$
** $p < .01$
* $p < .05$

Explaining fairness



Accuracy

"This is fair as the description is more accurate."

Physical visual characteristics

"I liked that it focused on aspects about the image, such as her hair and eye color."

Objective/Subjective

"This is more fair because it is not subjective and is accurate and less open to interpretation."

Understanding

"If someone gave me that I would be able to tell what the person looked like easier."

Political Correctness

"A lot of the words would not be described as favorable or putting the person in a good light."

Demographics

"It does not emphasize racial characteristics."

Background: FATE Community

- ACM Statement on Algorithmic Transparency and Accountability (detailed next) &
  Principles for Algorithmic Transparency and Accountability (7)
  - https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
- Industry
  - FATE groups, e.g., Microsoft https://www.microsoft.com/en-us/research/theme/fate/
- Academia
  - FAccT (formerly FAT) series of conferences https://facctconference.org/

ACM Principles

1. Awareness

Owners, designers, builders, users and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.

ACM Principles
2. Access and Redress

Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

ACM Principles
3. Accountability

Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.

ACM Principles
4. Explanation

Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.

ACM Principles
5. Data Provenance

A description of the way in which training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

ACM Principles
6. Auditability

Models, algorithms, data and decisions should be recorded so that they can be audited in cases where harm is suspected.

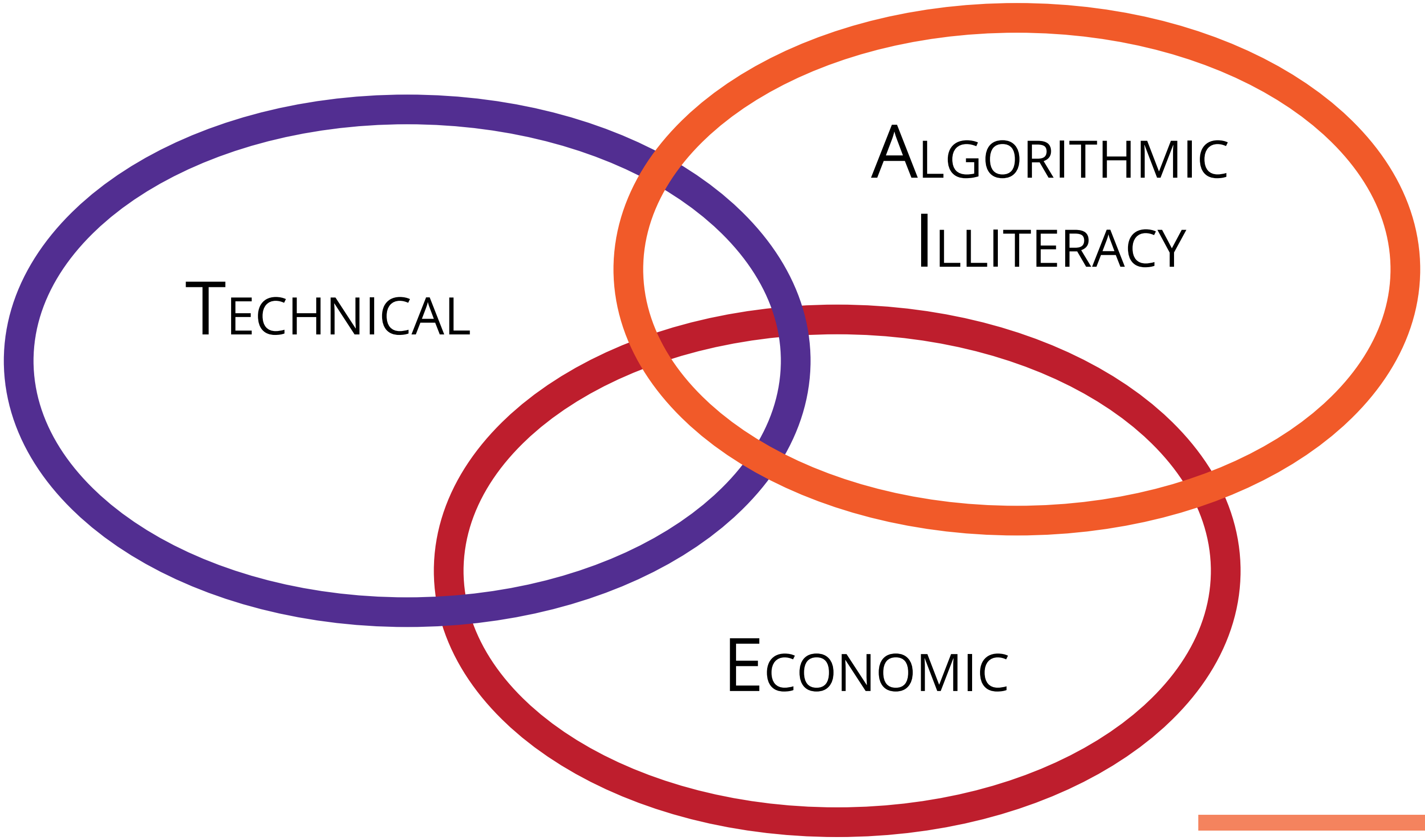
ACM Principles
7. Validation and Testing

Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

CHALLENGES

- What exactly does transparency mean?
- And fairness? Fair for whom?
  - 21 fairness definitions https://www.youtube.com/watch?v=jIXIuYdnyyk
- Bias - what is the baseline?
  - specific aspects of bias in ICT systems (e.g., based on age, gender, race, popularity, etc.)
- Diversity
  - different approaches and representations

Lack of Transparency



Algorithmic Illiteracy

Technical

Economic

Group Exercises

Awareness

Stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation and use and the potential harm that biases can cause to individuals and society.

Discover biases in **search engine results** and **auto-complete suggestions.** Design an experiment in which you test a number of queries, varying different parameters and examining the changes in the results. Parameters to try:

- Query language (e.g., Greek vs. English)
- Search engine used
  (e.g., Google vs. Bing vs. DuckDuckGo)
- Same search engine when you are identified or incognito
- Same search engine across users / members of the group

Awareness (2)

Variation:

You might also investigate the potential biases in the advertisements presented to users of search engines.

- Between members of the group (same query, same language)
- Varying the language of the query
- etc.

Access and Redress

Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

Explore the access and redress mechanisms for Facebook, Twitter, Instagram, and other social media.

Things to consider:

- Which mechanisms do they have in common and which are different?
- Can diverse sets of users (e.g., by age, region, level of digital literacy) exploit these mechanisms? Which challenges do you observe?
- Are their other measures and mechanisms that you would recommend?

Explanation

Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithms and the specific decision that are made.

Try to understand MovieLens (https://movielens.org) explanations on the movie recommendations. Sign in, define a profile, rate a few movies and check your suggested recommendations. Explain why they were suggested by MovieLens and elaborate on the reasons/facts as you understand them. Provide suggestions on improving their algorithm, and what else can be taken into consideration while creating explanations.

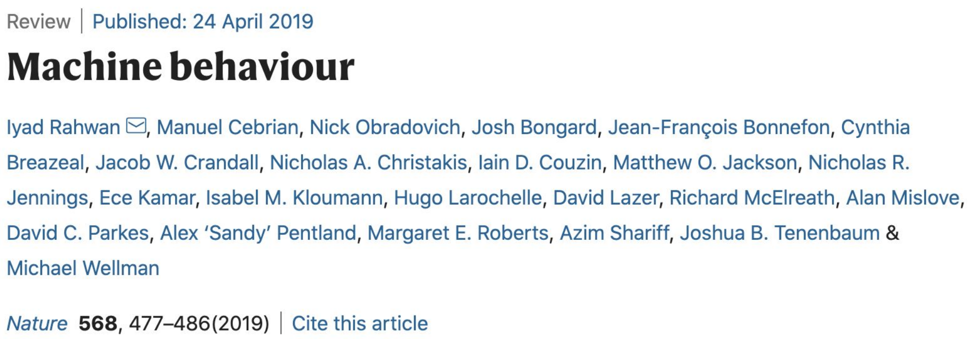
Explanation (2)

Variation:

You might also investigate explanations in other recommender systems that you use (e.g., Amazon, Netflix, etc.)

It is also interesting to compare explanations of the recommendations you receive over time, as your user profile evolves over time.

Discussion & Final Thoughts

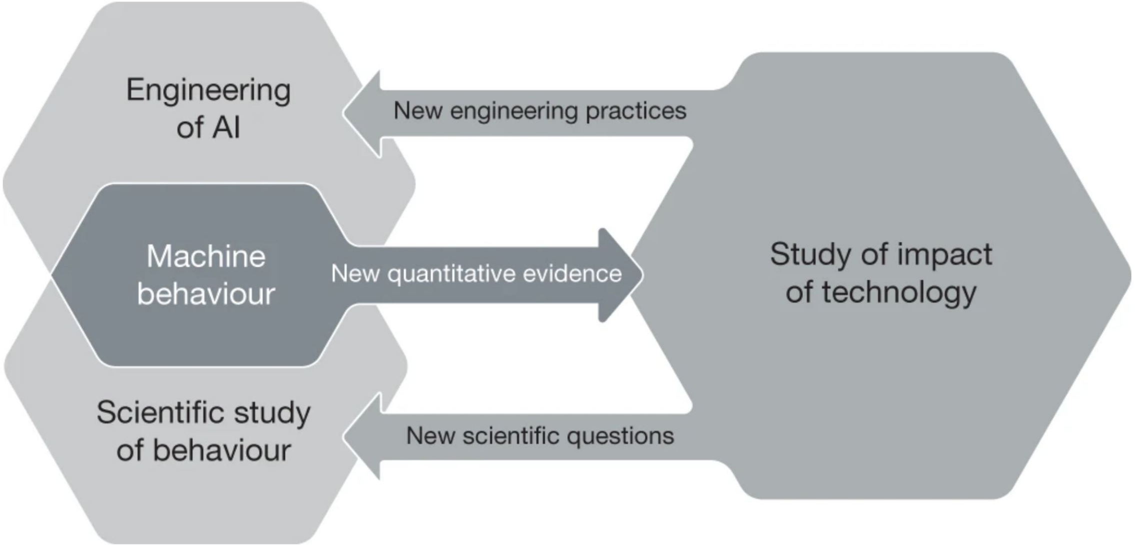


Review | Published: 24 April 2019

Machine behaviour

Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W. Crandall, Nicholas A. Christakis, Iain D. Couzin, Matthew O. Jackson, Nicholas R. Jennings, Ece Kamar, Isabel M. Kloumann, Hugo Larochelle, David Lazer, Richard McElreath, Alan Mislove, David C. Parkes, Alex ‘Sandy’ Pentland, Margaret E. Roberts, Azim Shariff, Joshua B. Tenenbaum & Michael Wellman

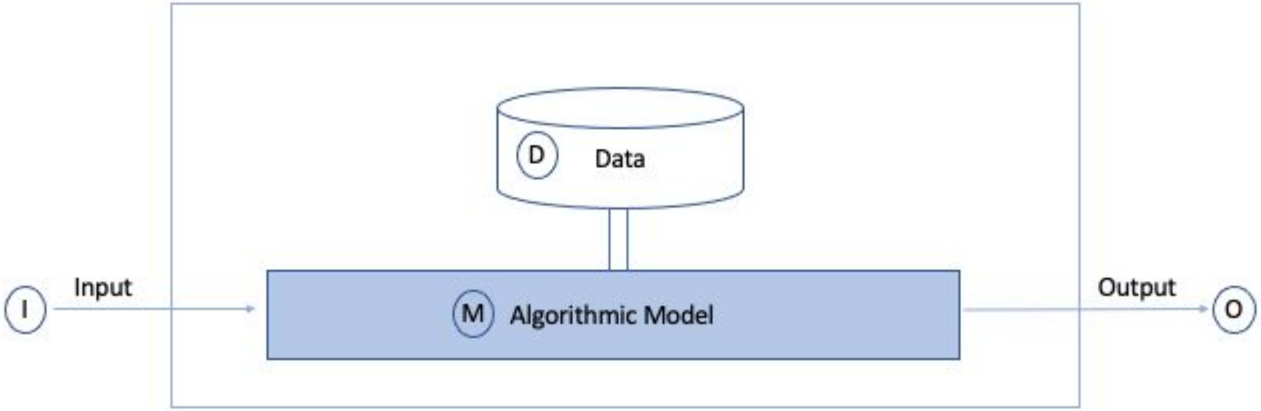
Nature **568**, 477–486(2019) | Cite this article

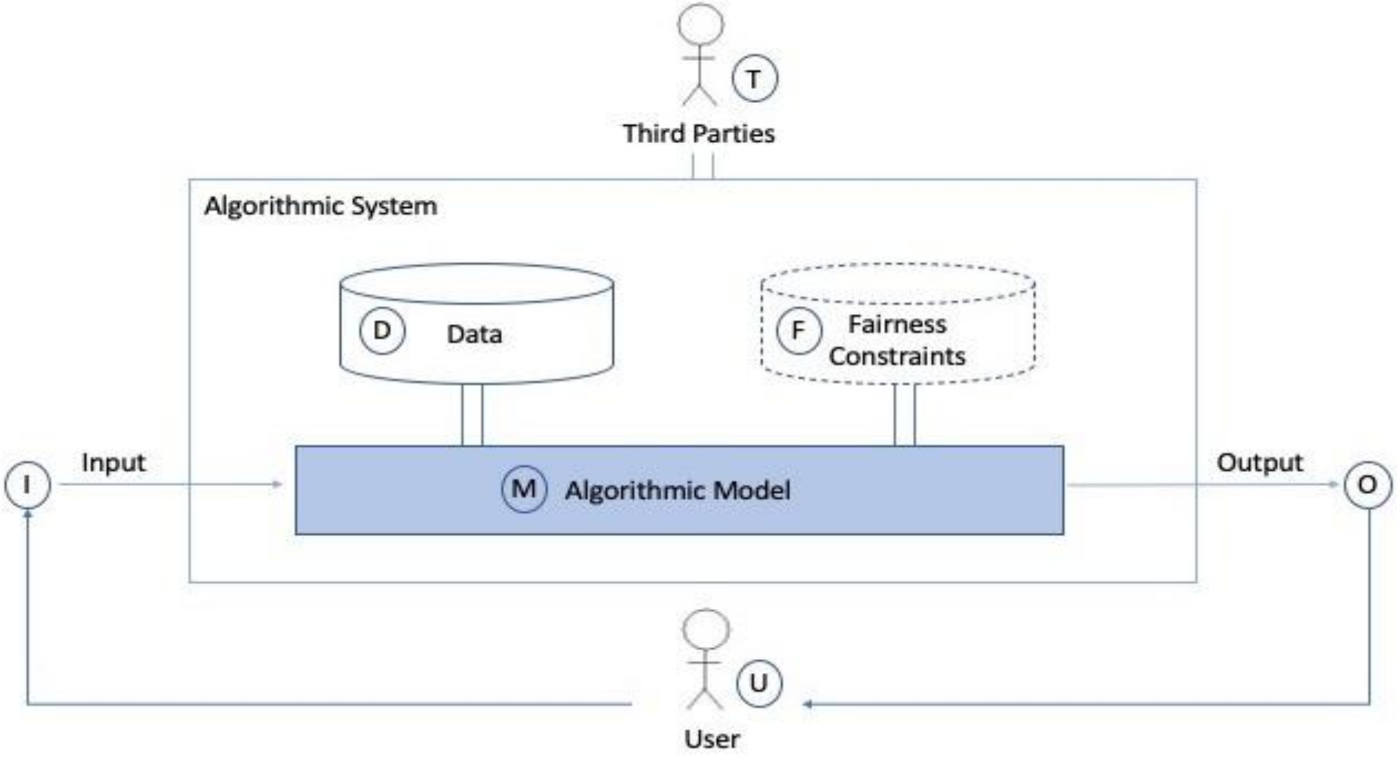

Engineering of AI
New engineering practices
Machine behaviour
New quantitative evidence
Study of impact of technology
Scientific study of behaviour
New scientific questions

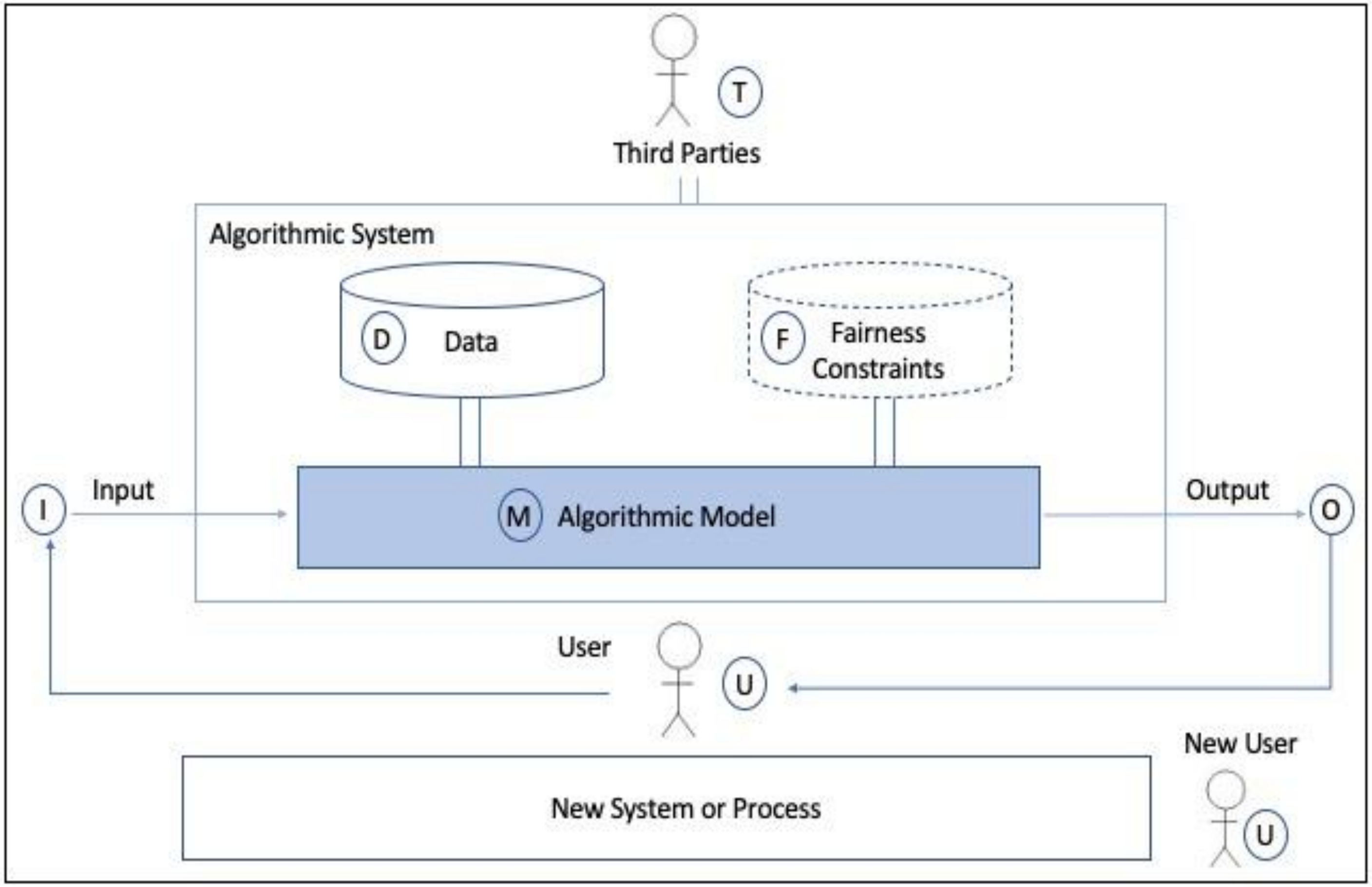
Thank you!

- www.**cycat.io**
- facebook.com/**CyCAT.EU**
- twitter.com/**CyCAT_EU**
- linkedin.com/in/**CyCAT**

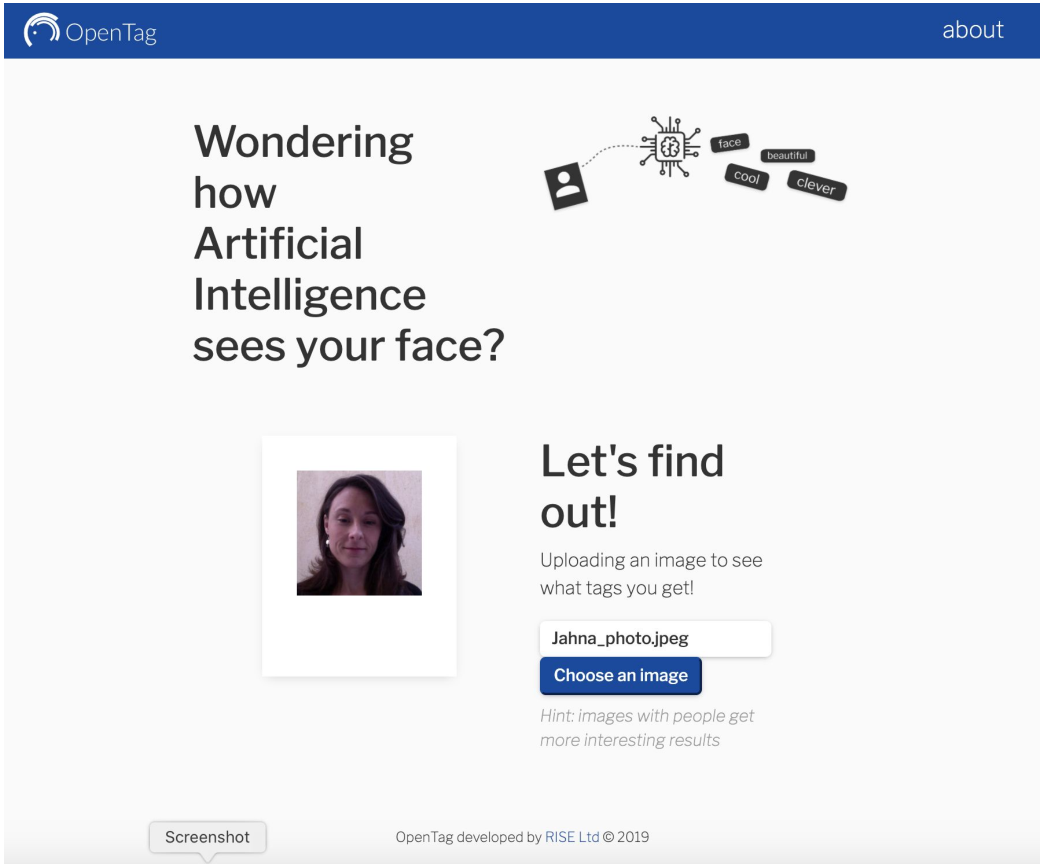
This research is partially funded by the European Union's Horizon 2020 research and innovation program under grant agreements No. 739578 (RISE), 810105 (CyCAT) and the Government of the Republic of Cyprus (RISE).


D
Data
I
Input
M
Algorithmic Model
Output
O


T
Third Parties
Algorithmic System
D
Data
F
Fairness
Constraints
I
Input
M
Algorithmic Model
Output
O
U
User


T
Third Parties
Algorithmic System
D
Data
F
Fairness Constraints
I
Input
M
Algorithmic Model
Output
O
User
U
New User
U
New System or Process

User Study – Invitation!



OpenTag

about

Wondering how Artificial Intelligence sees your face?

Let's find out!

Uploading an image to see what tags you get!

Jahna_photo.jpeg

Choose an image

Hint: images with people get more interesting results

Screenshot

OpenTag developed by RISE Ltd © 2019

http://ec2-34-255-198-84.eu-west-1.compute.amazonaws.com/opentag