

Artificial Intelligence #2

Lecture 11

Reveal and discuss the social and ethical issues existed in artificial intelligence

Lecturer YANG Lu

Agenda

- Existing challenges and Issues
 - AI/AS and job displacement issue
 - AI/AS and accountability issue
 - AI/AS and explainability/transparency/fairness issue
 - AI/AS governance
 - AI/Robots and human relationships

Issues Existed in Artificial Intelligence

Name a few:

- AI and job displacement
- Do humanlike machines deserve human rights?
- People might lost their sense of being unique
 - The boundary of human
- Human-robot relationship
- AI systems might be used towards undesirable ends
 - Massive surveillance program
- The use of AI systems might result into a loss of accountability, explainability, and transparency
- System bias in AI systems
- The boundary of AI
- etc.

AI/AI and Job Displacement

Automation and Job Displacement

Examples

Self-Driving Trucks

Tractor-trailers without a human at the wheel will soon barrel onto highways near you. What will this mean for the nation's 1.7 million truck drivers?



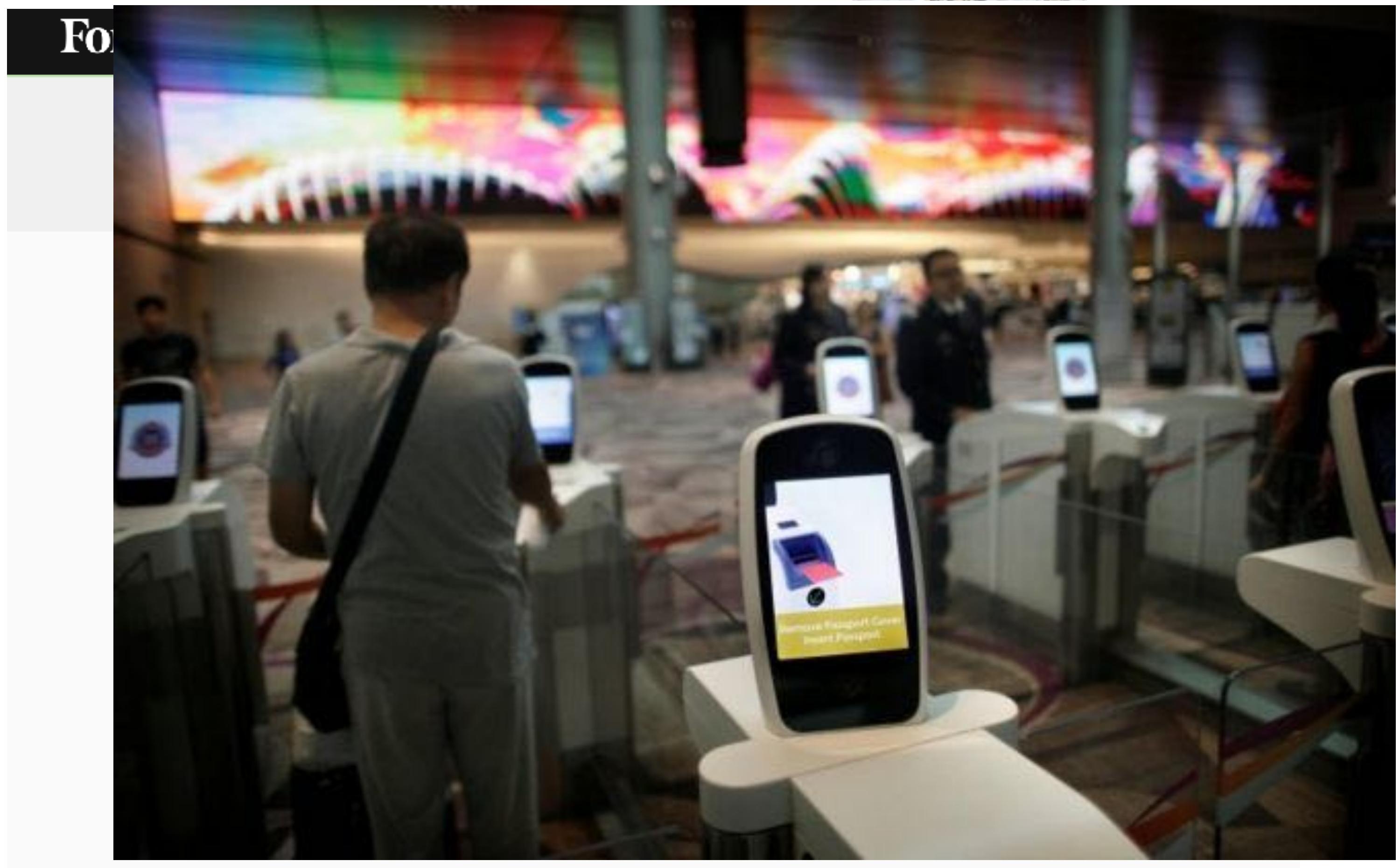
Drone Deliveries Are No Longer Pie In The Sky



Deutsche Post/DHL is testing deliveries of medicine from a pharmacy in Bonn. (Photo by Ulrich Baumgarten/Getty Images)

Automation and Job Displacement

More Automation Applications



Related - Technology and Job Displacement

South Korea cab

Drivers say the service being introduced is threatening their livelihoods

Last Published: Thu, Dec 20 2018, 07:32 PM

Bloomberg



Traffic jams were reported on some roads as drivers staged a strike Thursday, snarling traffic across the country. AFP

Seoul: Tens of thousands of drivers struck Thursday, snarling traffic as Uber-like ride-sharing services like Uber



Services like Uber and Lyft have won a victory over app-based ride-hailing services. The parliament resolved a measure to adjust to the changes.



01:35

Automation and Job Displacement

[“S’pore to be worst hit by job displacement as tech disrupts region’s labour markets: Study”, Sep 2018](#)

SINGAPORE — Workers in Singapore will be hit hard by technological disruptions with the Republic’s labour market set to face the largest degree of job displacement regionally in the next decade, based on a study on the impact of artificial intelligence (AI) on workers in six South-east Asian economies.

The new study by technology company Cisco and economic forecasting agency Oxford Economics also found that Singapore will have to confront the biggest mismatch between skills and jobs created among the countries in the region

This, said the team behind the study, is **due to Singapore’s digital transformation taking place at a faster rate than other Asean countries.**

IN NUMBERS

- By 2028, 28 million fewer workers in the region — more than 10 per cent of the present workforce across the six countries — will be needed to produce the same level of output as today.
- Agriculture is the industry forecast to be worst hit in Asean, accounting for 9.9 million full-time equivalent workers displaced.
- In Singapore, some 85,000 existing roles will disappear, pushing workers into other industries and vocations.
- The wholesale and retail sector is set to face the highest levels of displacement in Singapore by 2028 (80,000 jobs), followed by manufacturing (55,000) and transport (50,000).
- However, wholesale and retail is also poised to see the greatest job creation in Singapore in the same period (100,000 jobs). Manufacturing will come second (70,000 jobs), followed by finance and insurance (60,000 jobs).

Automation and Job Displacement

Solution?

Singapore SkillsFuture Program

SkillsFuture is a national movement to provide Singaporeans with the opportunities to develop their fullest potential throughout life, regardless of their starting points.

SkillsFuture Credit

All Singaporeans aged 25 and above will receive an opening credit of S\$500 from **January 2016**. Your credit will not expire and the government will provide periodic top-ups, so you may accumulate your credit.



JANUARY 14TH-20TH 2017

Lifelong learning

How to survive in the age of automation

A SPECIAL REPORT



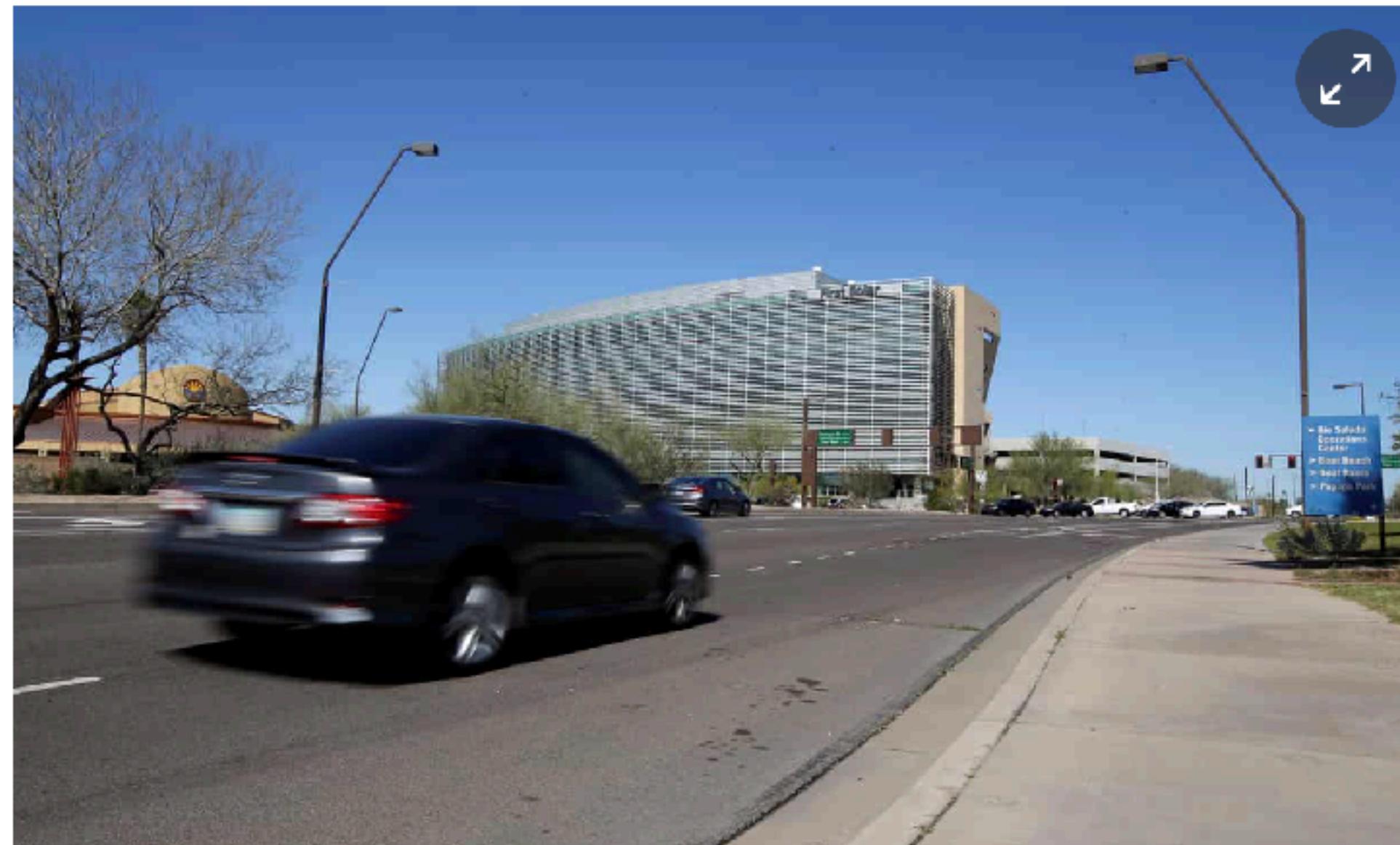
AI/AS and Accountability

Autonomous Vehicles and Accountability Issue

Self-driving Car Accidents

Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who later died at a hospital



A car passes the location where a woman pedestrian was struck and killed by an Uber self-driving sport utility vehicle in Tempe, Arizona, on Monday. Photograph: Rick Scuteri/Reuters

Tesla in fatal California crash was on Autopilot

31 March 2018

f Share



The driver of the Tesla Model X died shortly after the crash

REUTERS

Autonomous Vehicles and Accountability Issue

Whose fault?

Preliminary report on Uber's Arizona fatal case from NTSB:

"According to data obtained from the self-driving system, the system first registered radar and LIDAR observations of the pedestrian about 6 seconds before impact, when the vehicle was traveling at 43 mph. As the vehicle and pedestrian paths converged, **the self-driving system software classified the pedestrian as an unknown object, as a vehicle, and then as a bicycle with varying expectations of future travel path.** At 1.3 seconds before impact, the self-driving system determined that an emergency braking maneuver was needed to mitigate a collision (see figure 2).² According to Uber, **emergency braking maneuvers are not enabled while the vehicle is under computer control, to reduce the potential for erratic vehicle behavior. The vehicle operator is relied on to intervene and take action. The system is not designed to alert the operator.**"

<https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf>

Autonomous Vehicles and Accountability Issue

Whose fault?

Preliminary report on Tesla's California fatal case from NTSB:

- The Autopilot system was engaged on four separate occasions during the 32-minute trip, including a continuous operation for the last 18 minutes 55 seconds prior to the crash.
- During the 18-minute 55-second segment, the vehicle provided **two visual alerts and one auditory alert for the driver to place his hands on the steering wheel**. These alerts were made more than 15 minutes prior to the crash.
- During the 60 seconds prior to the crash, the driver's hands were detected on the steering wheel on three separate occasions, for a total of 34 seconds; for the last 6 seconds prior to the crash, the vehicle did not detect the driver's hands on the steering wheel.
- At 8 seconds prior to the crash, the Tesla was following a lead vehicle and was traveling about 65 mph.
- At 7 seconds prior to the crash, the Tesla began **a left steering movement while following a lead vehicle**.
- At 4 seconds prior to the crash, the Tesla was no longer following a lead vehicle.
- At 3 seconds prior to the crash and up to the time of impact with the crash attenuator, the Tesla's speed increased from 62 to 70.8 mph, **with no precrash braking or evasive steering movement detected**.
- <https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18FH011-preliminary.pdf>

Dilemma in Autonomous Vehicles Design

“If you delegate transportation from point A to Point B entirely to a machine, are you responsible if it hits someone?”

In a self-driving car, the control of the vehicle is shared between the driver and the car’s software. How the software behaves is in turn controlled—designed—by the software engineers.

In Aviation domain:

“while the plane is being run by software, the pilots in the cockpit are legally responsible for its operation. US Federal Aviation Administration (FAA) regulations specify this directly, and courts have consistently upheld it. So when something goes wrong, we observe pilots becoming “**moral crumple zones**”—largely totemic humans whose central role becomes soaking up fault, even if they had only partial control of the system.”

Case: Air France flight 447 in 2009

Automation Paradox:

the more efficient the automated system, the more crucial the human contribution of the operators. Humans are less involved, but their involvement becomes more critical.

Autonomous Vehicles and Accountability Issue

[“Everything you need to know about the Boeing 737 Max airplane crashes” - 22nd March, The Verge](#)

WHAT CAUSED THESE CRASHES?

Both crashes are currently under investigation, and there is no final word on what caused either tragedy. But investigators are focused on a specific tech feature that may have forced both planes into a nosedive seconds before the crashes.

A preliminary report from Indonesian investigators indicates that Lion Air 610 crashed because a faulty sensor erroneously reported that the airplane was stalling. The false report triggered [an automated system](#) known as the Maneuvering Characteristics Augmentation System, or MCAS. This system tried to point the aircraft's nose down so that it could gain enough speed to fly safely.

A FAULTY SENSOR ERRONEOUSLY REPORTED THAT THE AIRPLANE WAS STALLING

plane in an effort to prevent the plane from stalling.

Investigators have found strong similarities in the angle of attack data from both flights. A piece of a stabilizer in the wreckage of the Ethiopian jet with the trim set in an unusual position was similar to that of the Lion Air plane, *Reuters* reports.

MCAS takes readings from two sensors that determine how much the plane's nose is pointing up or down relative to oncoming airflow. When MCAS detects that the plane is pointing up at a dangerous angle, it can automatically push down the nose of the

WERE PILOTS GIVEN ADEQUATE TRAINING?

Short answer: no. When the Max jet was under development, regulators determined that pilots could fly the planes without extensive retraining because they were essentially the same as previous generations, [according to The New York Times](#). This saved Boeing a lot of money on extra training, which aided the company in its competition with Airbus to introduce newer, more fuel-efficient airplanes. The FAA didn't change those rules after Lion Air 610 crashed.

So rather than hours-long training sessions in giant, multimillion-dollar simulators, many pilots instead learned about the 737's new features on an iPad. Pilots at United Airlines put together a 13-page guide to the 737 Max, which did not mention the MCAS.

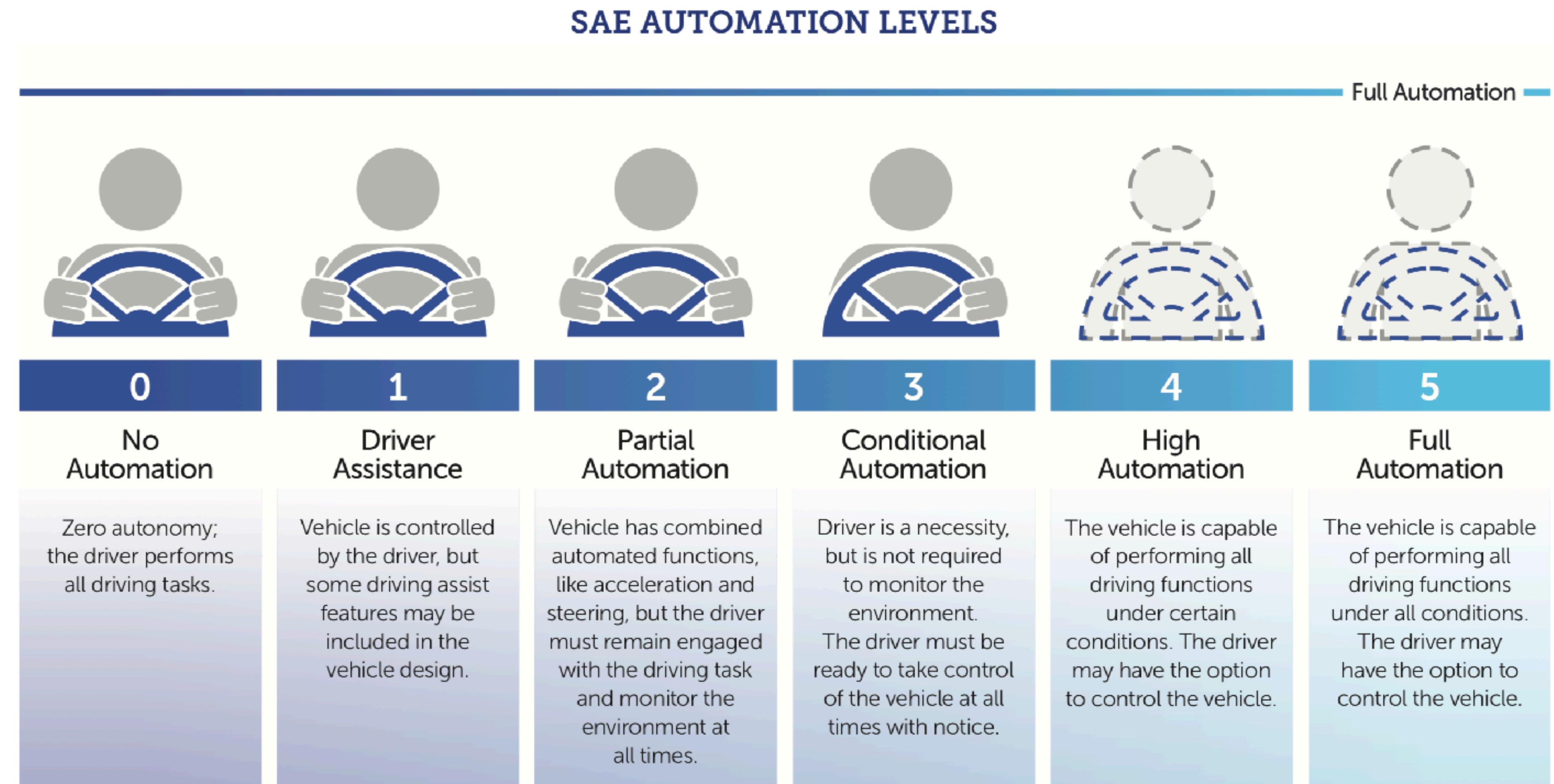
[According to Reuters](#), the doomed Lion Air cockpit voice recorder revealed how pilots scoured a manual in a losing battle to figure out why they were hurtling down to sea.

Since the crash of Ethiopian Airlines 302, that's mostly changed. On Sunday, March 17th, Muilenburg [issued a statement](#) describing steps the company was taking to update its technology. “While investigators continue to work to establish definitive conclusions, Boeing is finalizing its development of a previously-announced software update and pilot training revision that will address the MCAS flight control law’s behavior in response to erroneous sensor inputs,” Muilenburg said.

AS THE PLANE PLUMMETED, THE DOOMED LION AIR PILOTS SCOURED A MANUAL IN A LOSING BATTLE

Autonomous Vehicles Levels

Automation Levels (SAE: Society of Automotive Engineers)



Autonomous Vehicles and Regulations

Examples:

Singapore:

Revision on Road Traffic Act

Technical Reference (TR68), 2019

A set of national standards aimed at promoting the safe deployment of fully driverless vehicles in Singapore. Be more specific, to guide the industry on the development and deployment of vehicles in the Society of Automotive Engineers (SAE) Level 4 and 5 bands.

The guidelines mainly covers 4 dimensions:

- Vehicle behaviour
- Vehicle functional safety
- Cybersecurity
- Data formats

U.S.A.

Federal: H.R. 3888 - Self Drive Act (2017-2018)

e.g., privacy plan

Multiple states enacted relevant bills regarding autonomous vehicles regulations.

Safety rules/ standards regulatory gaps:

e.g., whether traditional controls, like a steering wheel, mirrors, and foot pedals still should be in an autonomous vehicles before it is allowed to operate on public roads?

e.g., whether human driver should always be there to take over the car in emergency?

e.g., metrics for human driver training

e.g., algorithm transparency

Autonomous Vehicles and Regulations

[2019 Autonomous Vehicles Readiness Index - KPMG](#)

Evaluation Dimensions

- Policy and legislation
- Technology and innovation
- Infrastructure
- Consumer acceptance

2 | Singapore



The government is positioning the city-state as a center for AV development with the deployment of a simulated urban test-bed and plans for driverless buses.

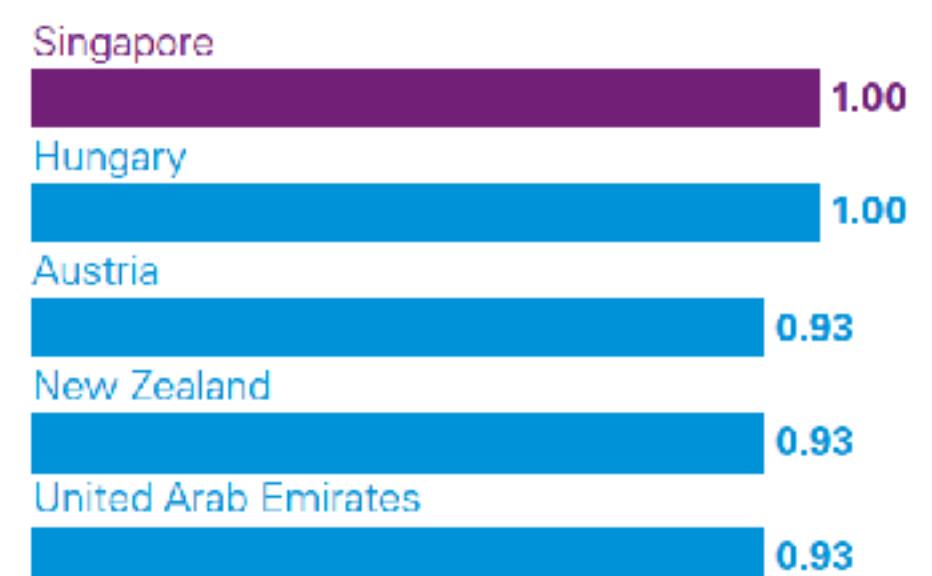
Singapore leads the policy and legislation and consumer acceptance pillars, and is second only to the Netherlands on infrastructure. Its government is working hard to

"The government is very proactive in thinking about the future of mobility. It is seriously investigating the possibilities as well as preparing for a regulatory environment that will facilitate a future that is autonomous," says Satya Ramamurthy, Partner, Head of Infrastructure, Government and Healthcare, KPMG in Singapore.

Singapore ranks relatively low on technology and innovation, at 15th of 25. Ramamurthy says Singapore's lack of conventional automotive manufacturing may prove to be a

21	n/a	Hungary	11.99
22	18	Russia	8.55
23	19	Mexico	7.73
24	20	India	6.87
25	17	Brazil	6.41

AV department within government – score by country (top 5)



Source: KPMG International, 2019

AI/AS and Explainability/Transparency/Fairness

Artificial Intelligence Race

Singapore Government Initiatives

AI Singapore (AISG)



A national AI programme launched by the National Research Foundation (NRF) to anchor deep national capabilities in Artificial Intelligence (AI) in 2017. NRF will invest up to \$150 million over five years in AI Singapore.

1. AI for everyone: free introductory AI program for 10,000 people ranging from secondary school students to working adults, over 3 years

2. AI for industry (AI4I): a three-month foundational programme for 2,000 working professionals and students who are technically inclined and eager to learn programming

3. AI Ethics Council: AI ethical conundrum - introduced by *having machines that make decisions on our behalf* – are a growing question which needs to be tackled and answered.

The formation of an Advisory Council on the Ethical Use of AI and Data

4. AI Apprenticeship Programme (AIAP): intensive 9-month programme

5. 100 Experiments (100E) Programme: aim to solve industrial problem by using AI technics

Artificial Intelligence Race

Singapore Government Initiatives

The full composition of Singapore's Advisory Council on the Ethical Use of AI and Data (Advisory Council) was announced by Minister for Communications and Information Mr S Iswaran at AI Singapore's first year anniversary.

SINGAPORE – 30 August, 2018: The full composition of Singapore's Advisory Council on the Ethical Use of AI and Data (Advisory Council) was today announced by Minister for Communications and Information Mr S Iswaran at AI Singapore's first year anniversary. This follows the earlier announcement in June on the establishment of the Advisory Council, to be chaired by former Attorney-General V.K. Rajah SC.

The eleven Advisory Council members come from diverse backgrounds and have been selected for their ability to contribute to the Advisory Council's objectives. Members comprise international leaders in AI such as Google, Microsoft and Alibaba; advocates of social and consumer interests; and leaders of local companies who are keen to make use of AI. The full list of Advisory Council members can be found in [Annex A](#).

The formation of the Advisory Council is one of three structured, interlinked initiatives¹ to support the engagement of stakeholders to collaboratively develop a trusted and vibrant AI ecosystem and position Singapore as a leading hub for AI.

The Advisory Council will advise and work with the Infocomm Media Development Authority (IMDA) on the responsible development and deployment of AI. Amongst other things, the Advisory Council will assist IMDA in engaging stakeholders on issues that support the development of AI governance capabilities and frameworks. These include engaging ethics boards of commercial enterprises on ethical and related issues arising from private sector use of AI and data; consumer representatives on consumer expectations and acceptance of the use of AI; as well as members of the private capital community on the need to incorporate ethical considerations in their investment decisions into businesses which develop or adopt AI.

The Advisory Council will also assist the Government in developing ethics standards and reference governance frameworks, and publish advisory guidelines, practical guides, and codes of practice for the voluntary adoption by the industry.

Supported by a secretariat, the Advisory Council will be able to tap on technical, legal, ethics and international experts from different fields and representatives for consumers and civil society to guide its work.

AI/AS Explainability/Transparency/Fairness

IMDA regulation on ethical use of AI and data

Two principles:

1) Decisions made by or with the help of AI are explainable, transparent and fair to consumers.

Explainability:

Ensure that automated and algorithmic decisions and any associated data driving those decisions can be explained to end-users and other stakeholders in non-technical terms.

Transparency:

Ensure that necessary information is disclosed and shared with relevant parties.

Fairness:

- a. Ensure that algorithmic decisions do not create discriminatory or unjust impacts across different demographic lines (e.g. race, sex, etc.).
- b. To develop and include monitoring and accounting mechanisms to avoid unintentional discrimination when implementing decision-making systems.
- c. To consult a diversity of voices and demographics when developing systems, applications and algorithms.

AI/AS Explainability/Transparency/Fairness

IMDA regulation on ethical use of AI and data

Two principles:

2) AI solutions are human-centric.

As AI is used to amplify human capabilities, the protection of the interests of human beings, including their well-being and safety, should be the primary considerations in the design, development and deployment of AI.

- a. To aim for an equitable distribution of the benefits of data practices and avoid data practices that disproportionately disadvantage vulnerable groups.
- b. To aim to create the greatest possible benefit from the use of data and advanced modelling techniques.
- c. Engage in data practices that encourage the practice of virtues that contribute to human flourishing, human dignity and human autonomy.
- d. To give weight to the considered judgments of people or communities affected by data practices and to be aligned with the values and ethical principles of the people or communities affected.
- e. To make decisions that should cause no foreseeable harm to the individual, or should at least minimise such harm (in necessary circumstances, when weighed against the greater good).
- f. To allow users to maintain control over the data being used, the context such data is being used in and the ability to modify that use and context.

AI/AS Explainability/Transparency/Fairness

NEW YORK CITY'S BOLD, FLAWED ATTEMPT TO MAKE ALGORITHMS ACCOUNTABLE

By Julia Powles December 20, 2017



Automated systems guide the allocation of everything from firehouses to food stamps. So why don't we know more about them?

Photograph by Mario Tama / Getty

The end of a politician's time in office often inspires a turn toward the existential, but few causes are as quixotic as the one chosen by James Vacca, who this month hits his three-term limit as a New York City Council member, representing the East Bronx. Vacca's nearly four decades in local government could well be defined by a bill that he introduced in August, and that passed last Monday by a unanimous vote. Once signed into law by Mayor Bill de Blasio, the legislation will establish a task force to examine the city's "automated decision systems"—the computerized algorithms that guide the allocation of everything from police officers and firehouses to public housing and food stamps—with an eye toward making them fairer and more open to scrutiny. In mid-October, I and some of my colleagues from a group at Cornell Tech that works on algorithmic accountability attended a hearing of the Council's technology committee to offer testimony on the bill. As Vacca, who chairs the committee, declared at the time, "If we're going to be governed by machines and algorithms and data, well, they better be transparent." Many of his constituents, he said, felt that "some inhuman computer is spitting them out and telling them where to go, and, if you don't like it, lump it."

Source: <https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable>

AI/AS Explainability/Transparency/Fairness

COMPAS Case

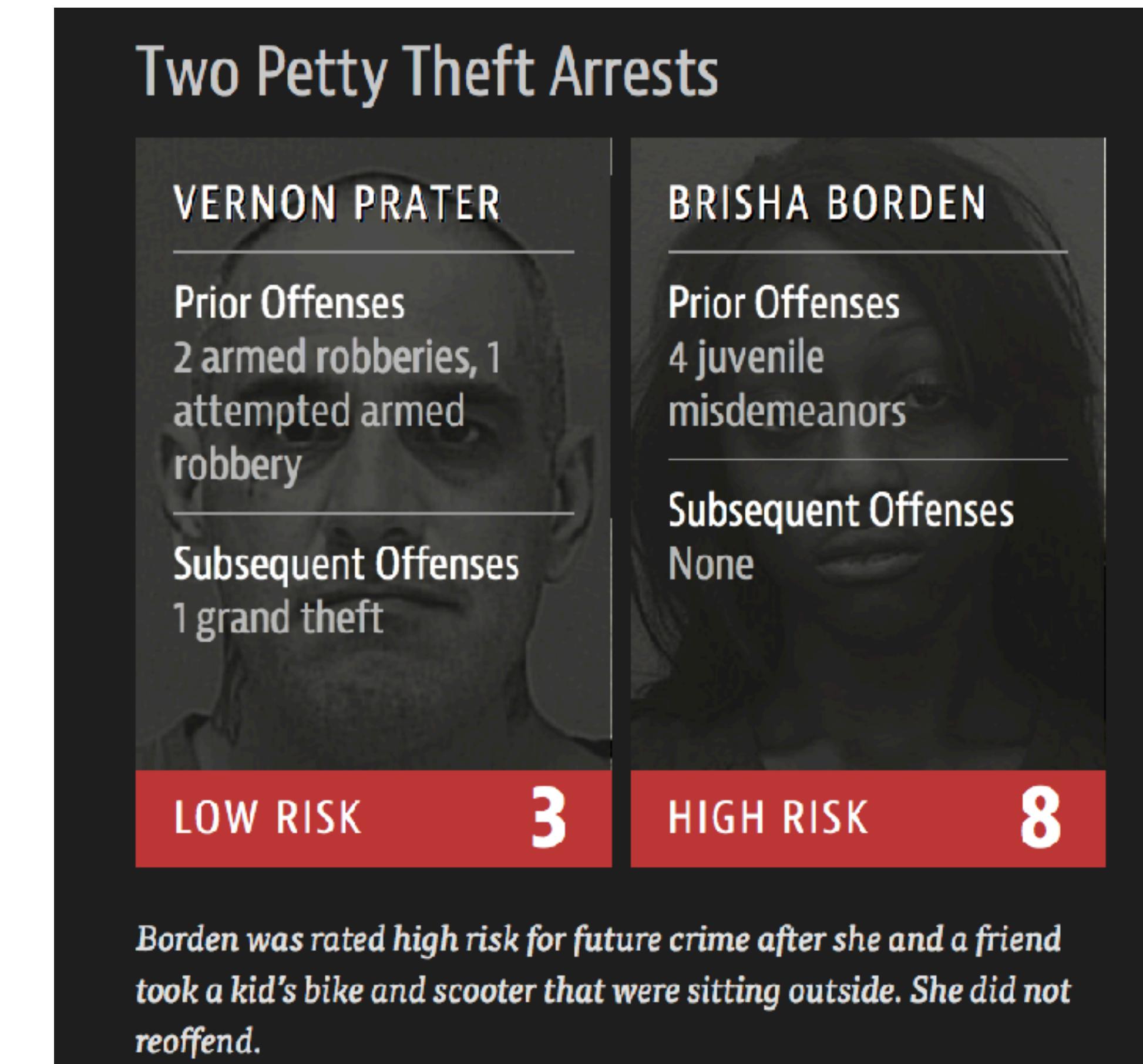
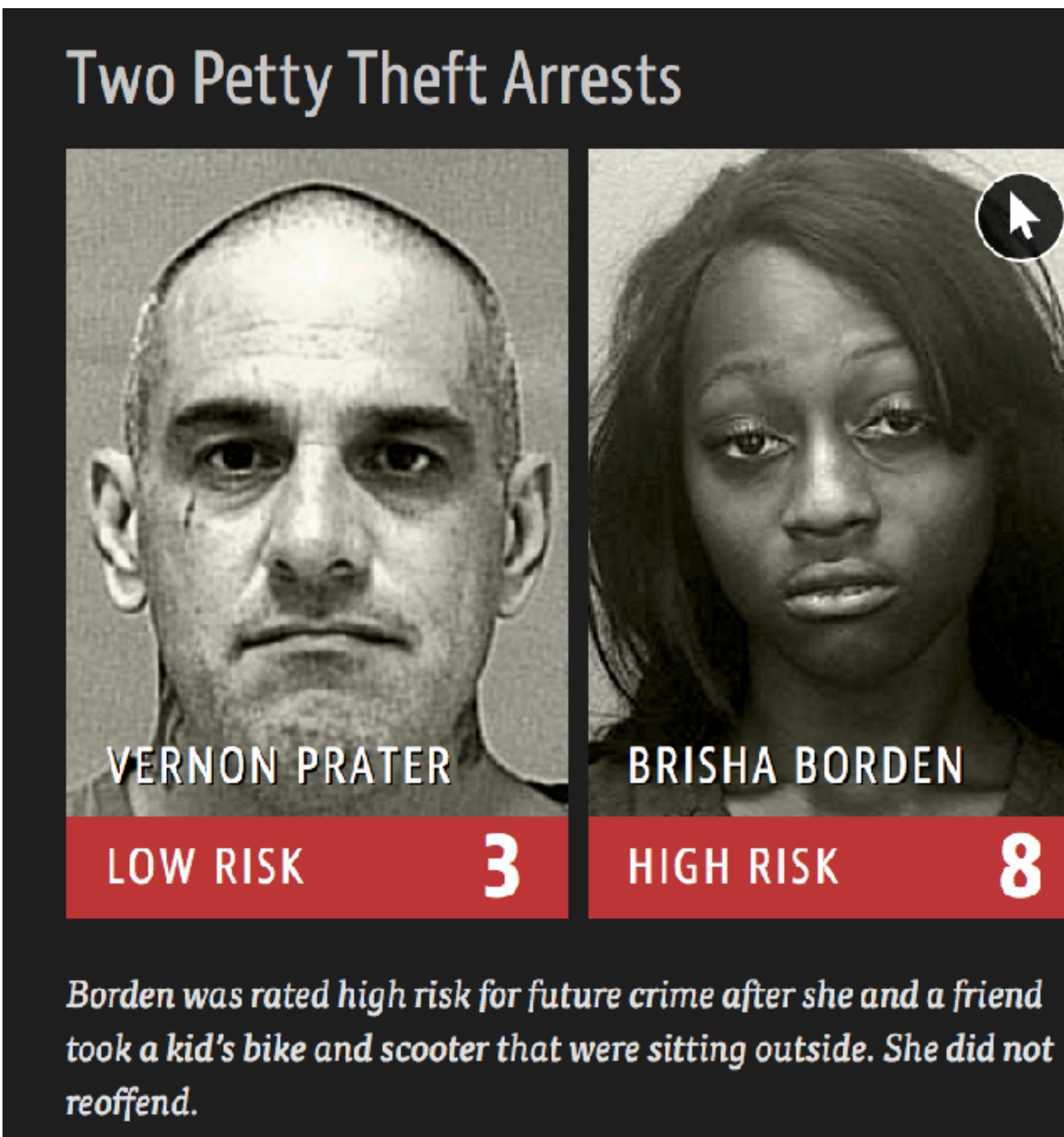
COMPAS (i.e., Correctional Offender Management Profiling for Alternative Sanctions) is an algorithm widely used in the US to guide sentencing by predicting the likelihood of a criminal reoffending. This software predicts a defendant's risk of committing a misdemeanor or felony within 2 years of assessment from **137 features** (not including race) about an individual and the individual's past criminal record.

In perhaps the most notorious case of AI prejudice, in May 2016 the US news organisation ProPublica reported that COMPAS is **racially biased**. According to the analysis, the system predicts that **black defendants pose a higher risk of recidivism than they do, and the reverse for white defendants**. Equivalent, the company that developed the software, disputes that. It is hard to discern the truth, or where any bias might come from, because the algorithm is **proprietary** and so not open to scrutiny.

Full report: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

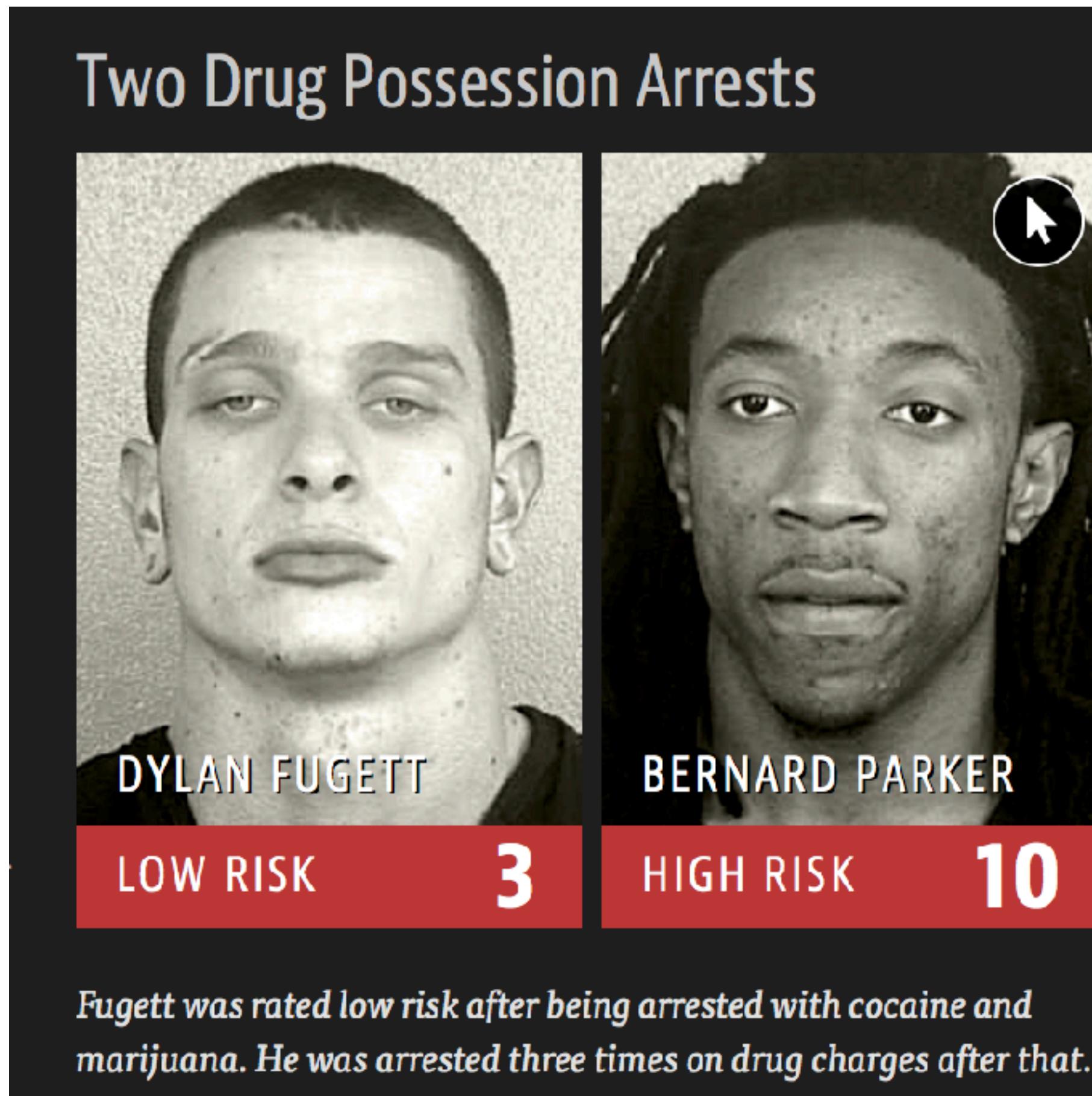
AI/AS Explainability/Transparency/Fairness

COMPAS Case : Two Petty Theft Arrests



AI/AS Explainability/Transparency/Fairness

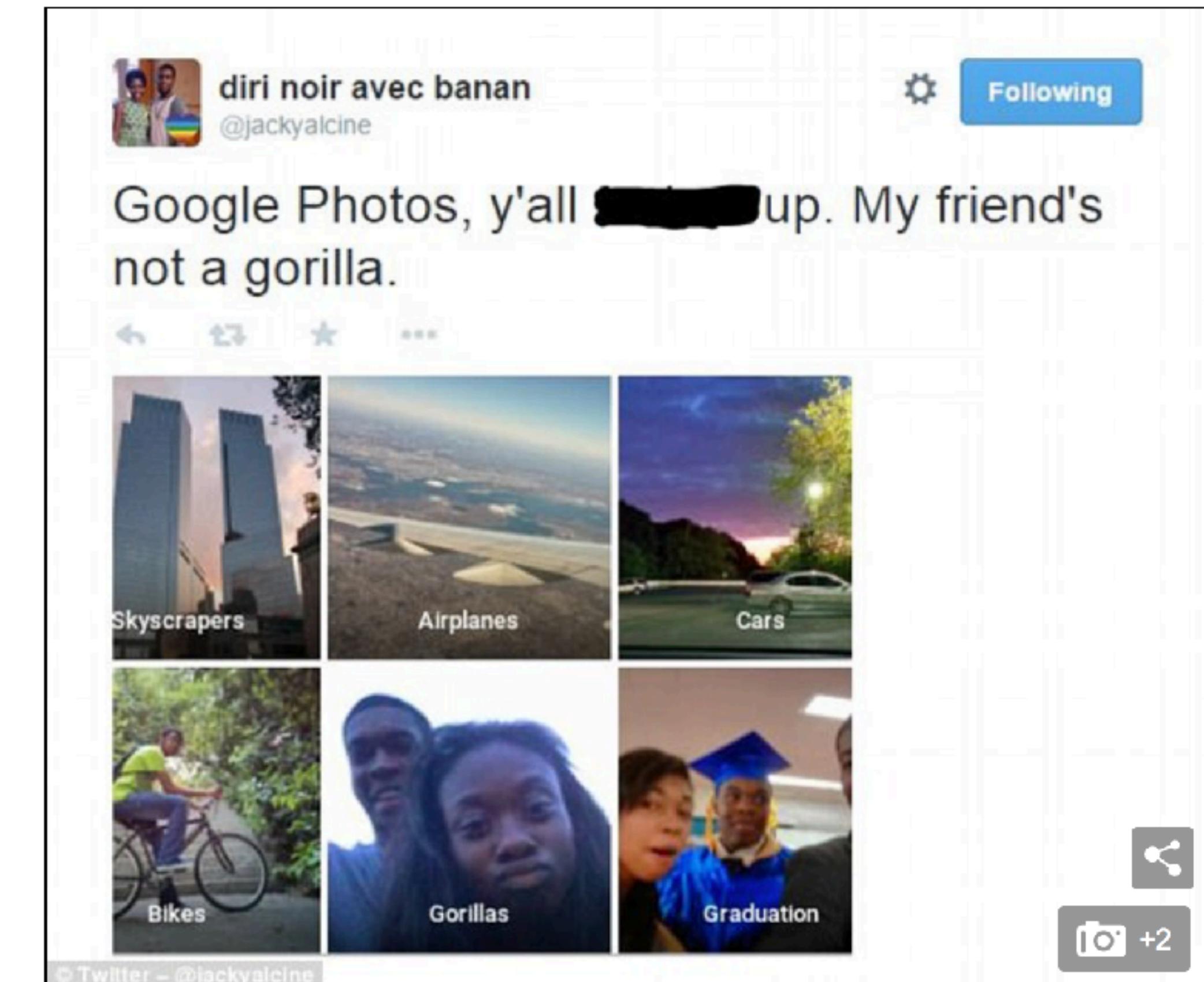
COMPAS Case : Two Drug Possession Arrests



Algorithmic Bias

Google is slammed for its attempt to 'fix' its racist image recognition algorithm by simply banning the term gorillas in its auto-tag tool after it labelled a black couple as apes

Source:<https://www.dailymail.co.uk/sciencetech/article-5270891/Google-bans-word-gorilla-racist-Photos-app.html>



In 2015 Jacky Alcine, from New York, spotted photographs of him and a female friend had been labelled as gorillas by Google Photos image recognition software. Google has admitted its image labelling technology is 'nowhere near perfect' but instead of fixing it the company has simply banned the term 'gorilla'

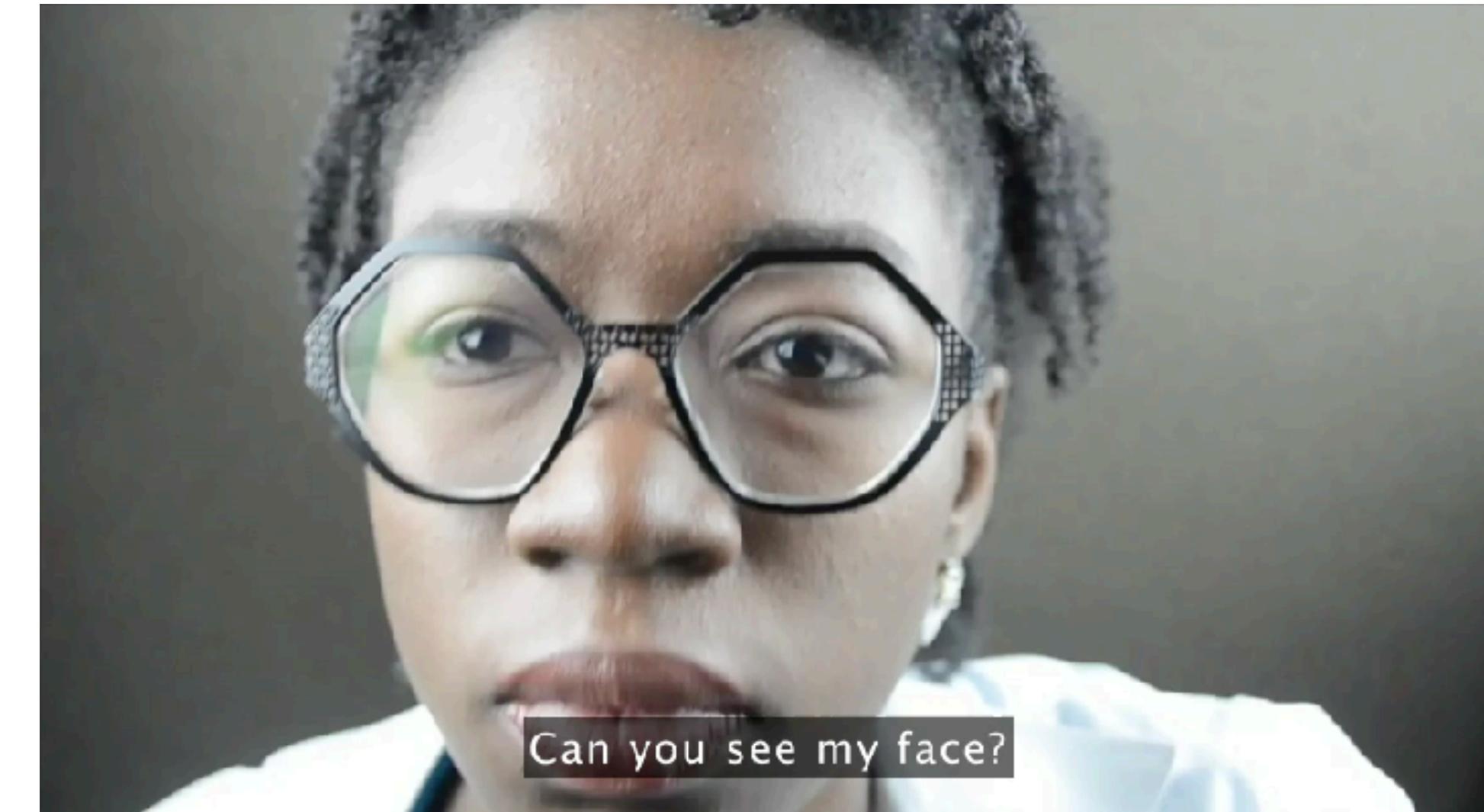
Algorithmic Bias

Computer systems can discriminate against individuals/groups of users

By denying opportunities or benefits, or assigns undesirable outcomes to particular users on grounds that are inappropriate

Selection bias issue:

This bias occurs when the data used to produce the model are not fully representative of the actual data or environment that the model may receive or function in.



Ted Talk: How I'm fighting bias in algorithms

[https://www.ted.com/talks/
joy_buolamwini_how_i_m_fighting_bias_in_algorithms?language=en](https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms?language=en)

AI Solutions and Human-Centric

[Google and Project Maven](#)



Google's New York office, in December. (Mark Lennihan/AP)

By [Drew Harwell](#)
June 1, 2018

Google will not seek to extend its contract next year with the Defense Department for artificial intelligence used to analyze drone video, squashing a controversial alliance that had raised alarms over the technological buildup between Silicon Valley and the military.

The tech giant will stop working on its piece of the military's AI endeavor known as Project Maven when its 18-month contract expires in March, a person familiar with Google's thinking told The Washington Post.

Diane Greene, the chief executive of Google's cloud-computing business, told employees of the decision at an internal meeting Friday, first [reported](#) by Gizmodo.

Google, which declined to comment, has faced widespread public backlash and employee resignations for helping develop technological tools that could aid in warfighting. The person said Google will soon release new company principles related to the ethical uses of AI.

Project Maven was [launched](#) in April 2017 to find ways the military could use AI to update its national security and defense capabilities "over increasingly capable adversaries and competitors," a Defense Department memo stated. In a pilot effort, AI was deployed to analyze hours of footage from Predator drones and other unmanned aircraft, pinpointing buildings and vehicles and processing video now tagged by human analysts.

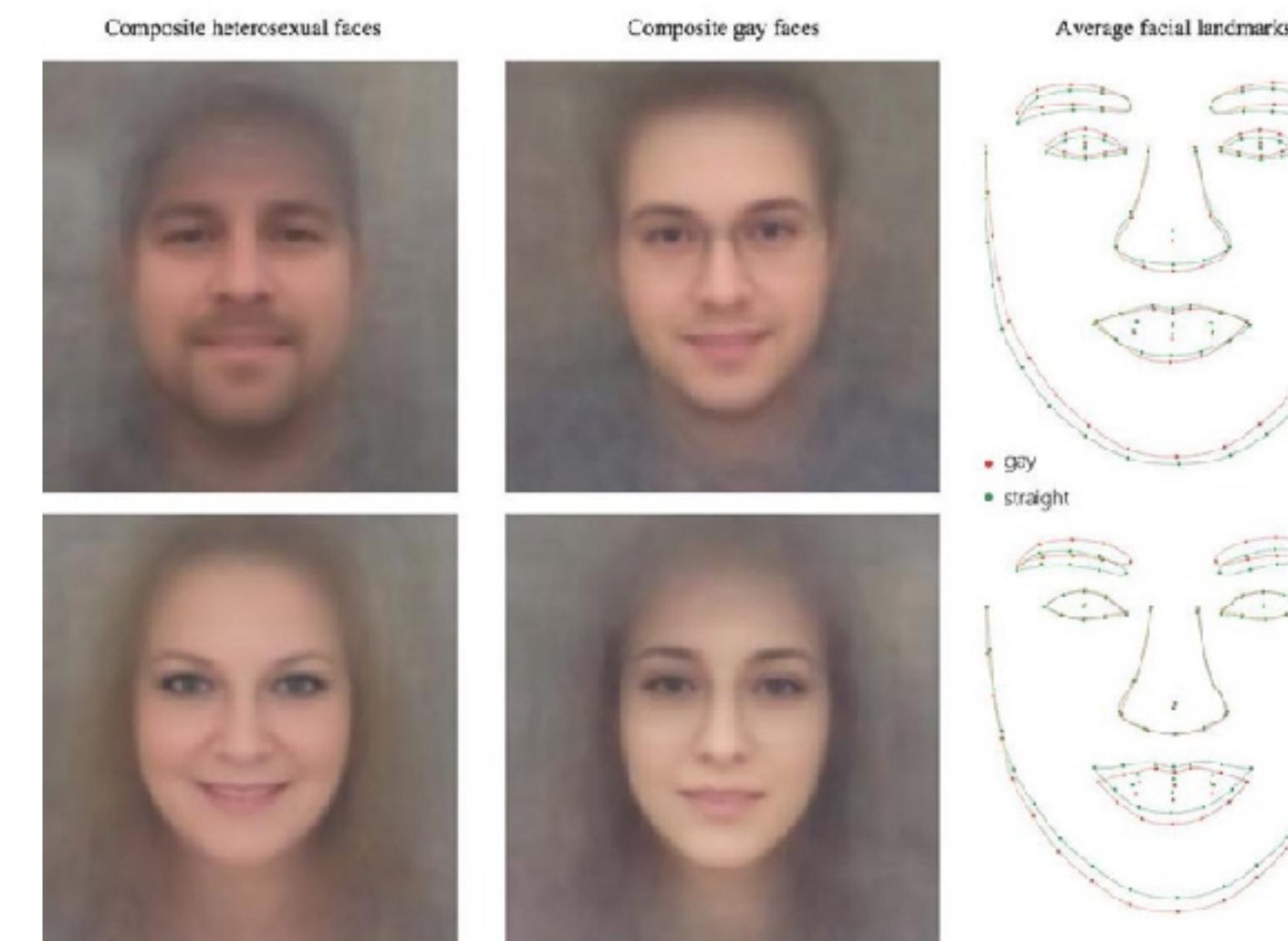
AI Solutions and Human-Centric

Gaydar Tool

 nytimes.com

Why Stanford Researchers Tried to Create a 'Gaydar' Machine

Michal Kosinski and Yilun Wang, co-authors of a



In one section of the study, the authors presented what they called “composite heterosexual faces,” left, and “composite gay faces,” right, built by averaging landmark locations of the faces classified as most and least likely to be gay. This did not sit well with some who suggested that this is just the latest example of physiognomy, rationalized by deep neural networks.

Michal Kosinski and Yilun Wang

Teaching a Machine to ‘See’ Sexuality

Dr. Kosinski and Mr. Wang began by copying, or “scraping,” photos from more than 75,000 online dating profiles of men and women in the United States. Those seeking same-sex partners were classified as gay; those seeking opposite-sex partners were assumed to be straight.

Some 300,000 images were whittled down to 35,000 that showed faces clearly and met certain criteria. All were white, the researchers said, because they could not find enough dating profiles of gay minorities to generate a statistically valid result.

AI Governance

A proposed model AI governance framework by PDPC, SG, Jan 2019

This Model Framework comprises guidance on measures promoting the responsible use of AI that organisations should adopt in the following key areas:

- a. **Internal Governance Structures and Measures:** Adapting existing or setting up internal governance structure and measures to incorporate values, risks, and responsibilities relating to algorithmic decision-making.
- b. **Determining AI Decision-Making Model:** A methodology to aid organisations in setting its risk appetite for use of AI, i.e. determining acceptable risks and identifying an appropriate decision-making model for implementing AI.
- c. **Operations Management:** Issues to be considered when developing, selecting and maintaining AI models, including data management.
- d. **Customer Relationship Management:** Strategies for communicating to consumers and customers, and the management of relationships with them.

AI Governance

A proposed model AI Governance Framework by PDPC, SG, Jan 2019

a. Internal Governance Structures and Measures:

- 1) Clear roles and responsibilities for the ethical deployment of AI
- 2) Risk management and internal controls

e.g., data inclusiveness, data bias check, review and monitoring system

b. Determining AI Decision-Making Model:

In determining the level of human oversight in an organisation's decision-making process involving AI, the organisation should consider the impact of such a decision on the individual using the probability-severity of harm matrix.

		Severity of Harm	
		High severity Low probability	High severity High probability
			Low severity Low probability
		High severity Low probability	Low severity High probability

Probability of Harm

AI Governance

A proposed model AI Governance Framework by PDPC, SG, Jan 2019

c. Operations Management

Data for AI model development

1) Understanding the lineage of data

knowing where the data originally came from, how it was collected, curated and moved within the organisation, and how its accuracy is maintained over time.

Keep Data Provenance Record

2) Ensuring data quality

e.g., accuracy, completeness, veracity of data (i.e., how credible the data is), recency, relevance, context of data collection, integrity of data (i.e., how well extraction and transformation have been performed), the usability of data, human interventions

3) Minimising inherent bias

e.g., selection bias, measurement bias, different dataset for model testing, training, and validation, periodic reviewing and updating of datasets.

“Where applicable, the model could also be checked for systematic bias by testing it on different demographic groups to observe whether any groups are being systematically advantaged or disadvantaged.”

Algorithm and Model

Explainability: challenges: IP, proprietary information,

Repeatability: the ability to consistently perform an action or make a decision, given the same scenario.

Traceability: audit trial, black box recorder

AI Governance

A proposed model AI Governance Framework by PDPC, SG, Jan 2019

d. Customer Relationship Management

1) general disclosure

Organisations should provide general information on whether AI is used in their products and/or services.

2) increased transparency

Organisations disclose the manner in which an AI decision may affect the individuals, and if the decision is reversible

3) easy-to-understand language

4) provide ethical evaluation and summary

5) policy for explanation

6) human-AI interface

7) option to opt-out

8) feedback and review channel

AI/Robotics and Humans

Robots Identity

On October 25, 2017, Sophia, a delicate looking woman with doe-brown eyes and long fluttery eyelashes made international headlines. She'd just become a full citizen of Saudi Arabia -- the first robot in the world to achieve such a status.

What does it mean to be a citizen?

What rights does Sophia hold?

Everything You Need To Know About Sophia, The World's First Robot Citizen



Human-Robots Relationship

China

Chinese man 'marries' robot he built himself

Zheng Jiajia had grown tired of pressure to get married so he turned to Yingying, a robot spouse he constructed last year



0

Benjamin Haas in Hong Kong

@haasbenjamin

Tuesday 4 April 2017
10.58 BST



Zheng Jiajia, 31, decided to 'marry' Yingying after failing to find a suitable human alternative. Photograph: Qiangjing Evening News

Source: <https://www.theguardian.com/world/2017/apr/04/chinese-man-marries-robot-built-himself>

25% of millennials think human-robot relationships will soon become the norm - study

Published time: 9 Dec, 2017 18:08

Edited time: 10 Dec, 2017 10:28

[Get short URL](#)



Stock images of a humanoid robot in Japan. © Reuters

A study of 12,000 people by French advertising firm Havas has revealed that human-robot relations could become ever more cozy in the years to come, with some 25 percent of people aged between 18-34 believing that it'll be normal for humans and robots to develop deep friendships, or even fall in love.

Issues to be Solved

Just list a few:

- How can we ensure that AI/AS (Autonomous Systems) do not infringe human rights (fair)?
- How can we assure that AI/AS are accountable?
- How can we ensure that AI/AS are transparent?
- How can we ensure that AI/AS are explainable?
- How can we extend the benefits and minimise the risks of AI/AS technology being misused?
- Lack of value-based ethical culture and practices for industry
- Lack of values-aware leadership
- Lack of ownership or responsibility from tech community
- Lack of empowerment to raise ethical concerns
- Lack of an independent review organization
- Lack of law regulation

2018 ACM A.M. Turing Award

The Association for Computing Machinery (ACM) awarded Yoshua Bengio, Geoffrey Hinton and Yann LeCun with what many consider the "Nobel Prize of computing," for the innovations they've made in AI.

"Working independently and together, Hinton, LeCun and Bengio *developed conceptual foundations for the field, identified surprising phenomena through experiments, and contributed engineering advances that demonstrated the practical advantages of deep neural networks.*"

source: <https://www.acm.org/media-center/2019/march/turing-award-2018>

5,402 views | Mar 27, 2019, 07:41pm

Turing Award And \$1 Million Given To 3 AI Pioneers



Winners of Turning Award NEW YORK TIMES

Recommended Materials

Are Robots Competing for Your Job? The New Yorker, March, 2019, <https://www.newyorker.com/magazine/2019/03/04/are-robots-competing-for-your-job>

How to Survive in the Age of Automation, A Special Report, The Economist, Jan, 2017

2019 Autonomous Vehicles Readiness Index - KPMG, <https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/02/2019-autonomous-vehicles-readiness-index.pdf>

A PROPOSED MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK, Jan 2019, PDPC Singapore, <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/A-Proposed-Model-AI-Governance-Framework-January-2019.pdf>

Everything You Need To Know About Sophia, The World's First Robot Citizen, Nov 2017, Forbes, <https://www.forbes.com/sites/zarastone/2017/11/07/everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/#8a5df8e46fa1>

Recommended Materials

“Whose Life Should Your Car Save?” by Azim Shariff et al. (New York Times, 2016).

“Executive Summary: The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems” (IEEE).

“Technology is Biased too. How do we Fix it?” By Laura Hudson (Five Thirty Eight, July 20, 2017)

“Data Monopolists Like Google Are Threatening the Economy” by Kira Radinsky (HBR, 2015)

“Transparent, explainable, and accountable AI for robotics” by Sandra Wachter, Brent Mittelstadt, and Luciano Floridi (Science, 2017).

“Is Effective Regulation of AI Possible? Eight Potential Regulatory Problems” by John Danaher (Philosophical Disquisition 2015)

“The Relentless Pace of Automation” by David Rotman (MIT Technology Review, 2017)

The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford #NIPS2017, https://www.youtube.com/watch?v=fMym_BKWQzk

Thank You