



**UNSW**  
SYDNEY

# Ethics in Computer Science **COMP4920**

Week 5: Ethics Framework, Accountability, and Privacy  
Flora Salim

# Acknowledgment of Country

I would like to acknowledge the Bedegal people that are the Traditional Custodians of this land. I would also like to pay my respects to the Elders both past and present and extend that respect to other Aboriginal and Torres Strait Islanders who are present here today.

# Agenda

- Value Sensitive Design
- AI ethics
- Reliability and Accountability
- Data Privacy and Governance
- Fairness and Transparency (a quick glance)

# Value Sensitive Design (VSD)

perspective

## *Value-Sensitive Design*

Batya Friedman Colby College and The Mina Institute

Values emerge from the tools that we build and how we choose to use them. Yet, in most of the current practice in designing computer technology and the related infrastructure of cyberspace, little is said about values.

VSD:

*An interactional theory and method that accounts for human values in a principled and structured manner throughout the design process*

(Friedman 1996; Friedman, Hendry, and Borning 2017)

Friedman B. Value-sensitive design. interactions. 1996 Dec 1;3(6):16-23.



UNSW  
SYDNEY

# Human Values (Friedman, Kahn, Borning, Hultgren, 2013)

Table 4.1 Human values (with Ethical Import) often implicated in system design

From: [Value Sensitive Design and Information Systems](#)

Human value	Definition	Sample literature
Human welfare	Refers to people's physical, material, and psychological well-being	Leveson (1991), Friedman et al. (2003), Neumann (1995), Turiel (1983, 1998)
Ownership and property	Refers to a right to possess an object (or information), use it, manage it, derive income from it, and bequeath it	Becker (1977), Friedman (1997b), Herskovits (1952), Lipinski and Britz (2000)
Privacy	Refers to a claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others	Agre and Rotenberg (1998), Bellotti (1998), Boyle et al. (2000), Friedman (1997b), Fuchs (1999), Jancke et al. (2001), Palen and Dourish (2003), Nissenbaum (1998), Phillips (1998), Schoeman (1984), Svensson et al. (2001)
Freedom from bias	Refers to systematic unfairness perpetrated on individuals or groups, including pre-existing social bias, technical bias, and emergent social bias	Friedman and Nissenbaum (1996), cf. Nass and Gong (2000), Reeves and Nass (1996)
Universal usability	Refers to making all people successful users of information technology	Aberg and Shahmehri (2001), Shneiderman (1999, 2000), Cooper and Rejmer (2001), Jacko et al. (1999), Stephanidis (2001)
Trust	Refers to expectations that exist between people who can experience good will, extend good will toward others, feel vulnerable, and experience betrayal	Baier (1986), Camp (2000), Dieberger et al. (2001), Egger (2000), Fogg and Tseng (1999), Friedman et al. (2000a), Kahn and Turiel (1988), Mayer et al. (1995), Olson and Olson (2000), Nissenbaum (2001), Rocco (1998)
Autonomy	Refers to people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals	Friedman and Nissenbaum (1997), Hill (1991), Isaacs et al. (1996), Suchman (1994), Winograd (1994)
Informed consent	Refers to garnering people's agreement, encompassing criteria of disclosure and comprehension (for "informed") and voluntariness, competence, and agreement (for "consent")	Faden and Beauchamp (1986), Friedman et al. (2000b), The Belmont Report (1978)
Accountability	Refers to the properties that ensures that the actions of a person, people, or institution may be traced uniquely to the person, people, or institution	Friedman and Kahn (1992), Friedman and Millet (1995), Reeves and Nass (1996)
Courtesy	Refers to treating people with politeness and consideration	Bennett and Delatre (1978), Wynne and Ryan (1993)
Identity	Refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time	Bers et al. (2001), Rosenberg (1997), Schiano and White (1998), Turkle (1996)
Calmness	Refers to a peaceful and composed psychological state	Friedman and Kahn (2003), Weiser and Brown (1997)
Environmental sustainability	Refers to sustaining ecosystems such that they meet the needs of the present without compromising future generations	United Nations (1992), World Commission on Environment and Development (1987), Hart (1999), Moldan et al. (1997), Northwest Environment Watch (2002)

# Value Sensitive Design (VSD)

VSD: “An approach that rigorously accounts for human values in the technical design and engineering process”

VSD projects:

- begins with the identification of stakeholders
- surfacing of their values through conceptual and empirical investigations

Stephanie Ballard, Karen M. Chappell, and Kristen Kennedy. 2019. Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. In Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19). Association for Computing Machinery, New York, NY, USA, 421- 433. <https://doi.org/10.1145/3322276.3323697>



# Value-Sensitive Design



Conceptual  
Investigations



Empirical  
Investigations



Technical  
Investigations



UNSW  
SYDNEY

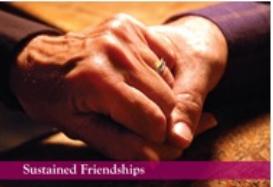
# A VSD toolkit: Envisioning cards

# Envisioning Cards

Home    About    Get the Cards    Ideas for Use    Contact the Authors

## The Envisioning Criteria

The Envisioning Cards are built upon a set of five Envisioning Criteria that are intended to raise awareness of long-term and systemic issues in design. Each Envisioning Card represents a specific theme within one of these Envisioning Criteria. Based on roughly two decades of research into accounting for human values in the design of technology, the Envisioning Cards were developed by the Value Sensitive Design Research Lab at the Information School at the University of Washington.

Stakeholders	Time	Values	Pervasiveness	Multi-lifespan
 Indirect Stakeholders	 Sustained Friendships	 Value Tensions	 Crossing National Boundaries	 Material Longevity
Drawing from Value Sensitive Design methodology, the Stakeholder criterion emphasizes the range of effects of a technology, both on those who are in direct contact with a technology (direct stakeholders), and on those who might not be direct users, but whose lives are nevertheless affected by various interactions around the technology (indirect stakeholders).	Inspired by the long-term perspective of urban planning, the Time criterion helps guide designers to consider the longer term implications of their work – implications that will only emerge after the technology has moved through initial phases of novelty to later phases of appropriation and integration into society.	The Value criterion emphasizes the impact of technology on human values. Our use of the term values draws from the Value Sensitive Design literature, "what a person or group of people consider important in life." In interaction design, we have found values of interest to include but are not limited to: autonomy, community, cooperation, democratization, environmental sustainability, expression, fairness, human dignity, inclusivity and exclusivity, informed consent, justice, ownership, privacy, self-efficacy, security, trust, and universal access.	The Pervasiveness criterion emphasizes systemic interactions that follow from the widespread adoption of an interactive technology. Technologies can become pervasive with respect to geographic (e.g., city navigation software use within urban areas), cultural (e.g., text messaging within the deaf community), demographic (e.g., online social networking sites among teenagers), and other factors.	Multi-lifespan design is primarily concerned with significant societal problems that defy rapid solution. Genocide, HIV/AIDS, famine, deforestation, species extinction, forced exodus – these problems share some commonalities. In one way or another, they entail widespread losses to human beings, to other sentient beings, or to the natural world. The structure of these problems and their solution spaces require long periods of time to unfold. In turn, these problems call for designers to take a more proactive and long term approach to how we envision and design technology. The Multi-lifespan criterion prompts consideration and discussion of such concerns within the context of designing tools, technology, policy, and infrastructure.

[https://www.envisioningcards.com/?page\\_id=2](https://www.envisioningcards.com/?page_id=2)

# SafetyNet: a hypothetical scenario

*SafetyNet is a hypothetical, commercial software platform which leverages publicly available demographic and criminal data, mapping technology, and satellite-tracking capabilities to create maps for display on various mobile technologies (eg cell phones, blackberries, in-car navigational systems). These maps are used to alert urban travelers as they venture into potentially unpleasant or dangerous areas during their travel.*

Taken from:

Nathan LP, Klasnja PV, Friedman B. Value scenarios: a technique for envisioning systemic effects of new technologies. In CHI'07 extended abstracts on Human factors in computing systems 2007 Apr 28 (pp. 2585-2590)



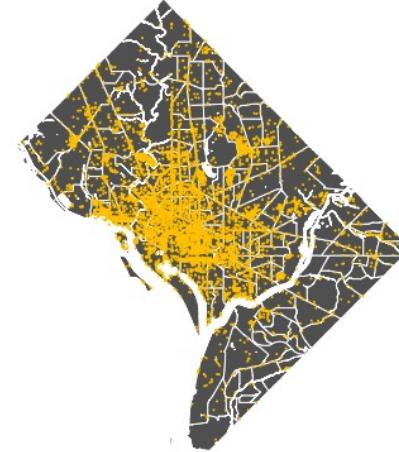
# Case study: Pokemon Go

[https://www.urban.org/  
urban-wire/pokemon-  
go-changing-how-cities-  
use-public-space-could-  
it-be-more-inclusive](https://www.urban.org/urban-wire/pokemon-go-changing-how-cities-use-public-space-could-it-be-more-inclusive)

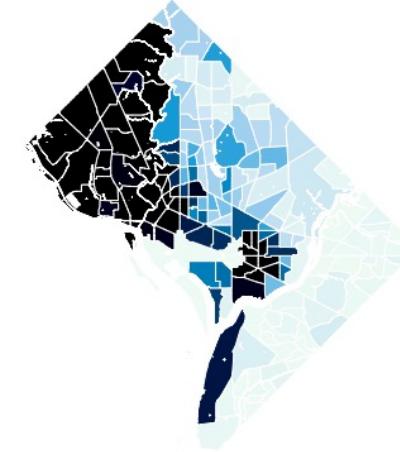
Ingress Portals Are Concentrated in Majority-White Census Tracts



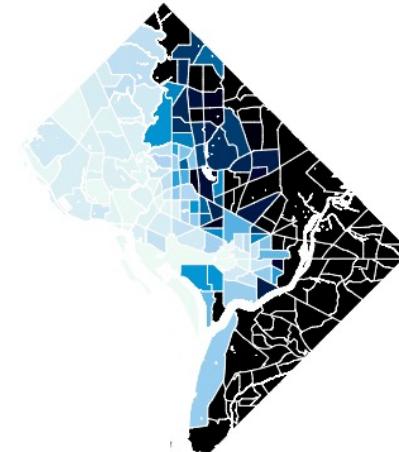
Portal locations



Share white, non-Hispanic



Share black, non-Hispanic



Share ages 18-29



Sources: Ingress Intel Map; American Community Survey 2010-14.

URBAN INSTITUTE



UNSW  
SYDNEY

# Judgment Call: The Game



Judgment Call is an award-winning game and team-based activity that puts Microsoft's AI principles of fairness, privacy and security, reliability and safety, transparency, inclusion, and accountability into action. The game provides an easy-to-use method for cultivating stakeholder empathy by imagining their scenarios. Game participants write product reviews from the perspective of a particular stakeholder, describing what kind of impact and harms the technology could produce from their point of view.



# ACM Principles for Algorithmic Transparency and Accountability

---

1. Awareness

---

2. Access and redress

---

3. Accountability

---

4. Explanation

---

5. Data Provenance

---

6. Auditability

---

7. Validation and Testing

[https://www.acm.org/binaries/  
content/assets/public-  
policy/2017\\_usacm\\_statement  
\\_algorithms.pdf](https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf)

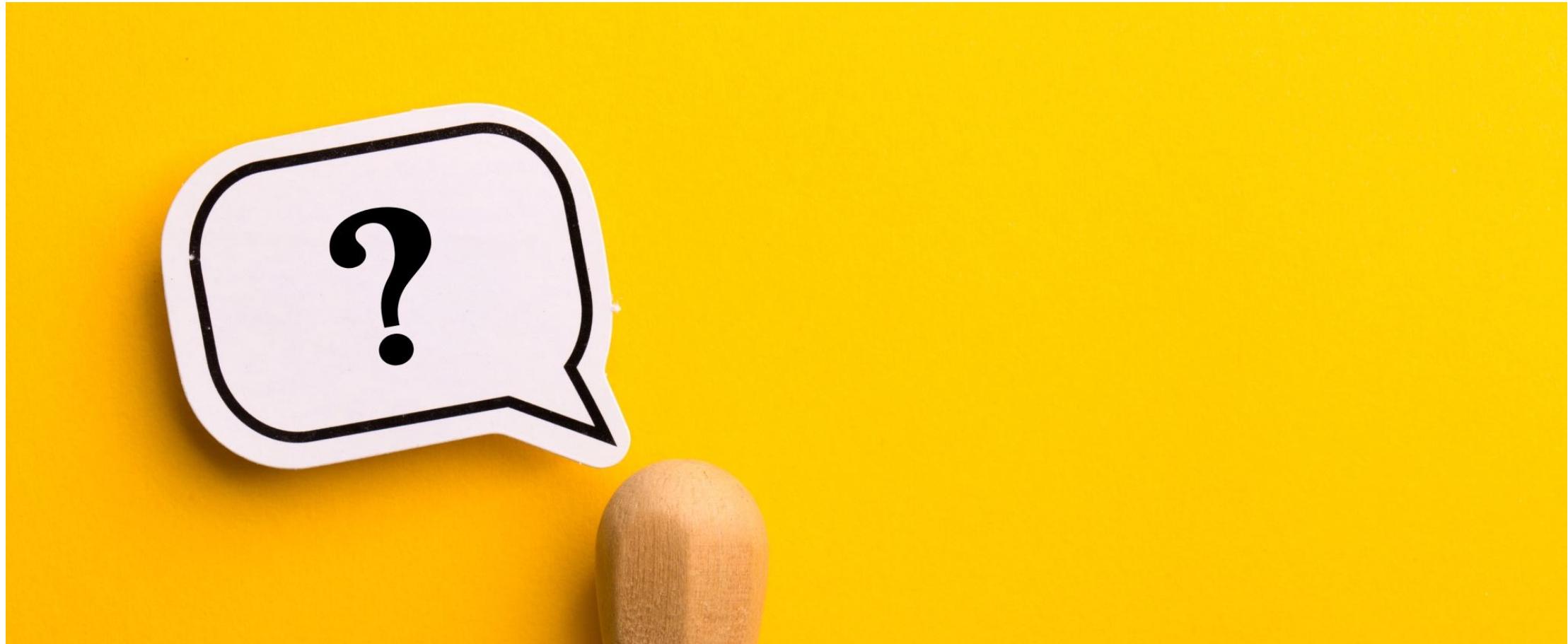


# AI Ethics



**UNSW**  
SYDNEY

# Why AI

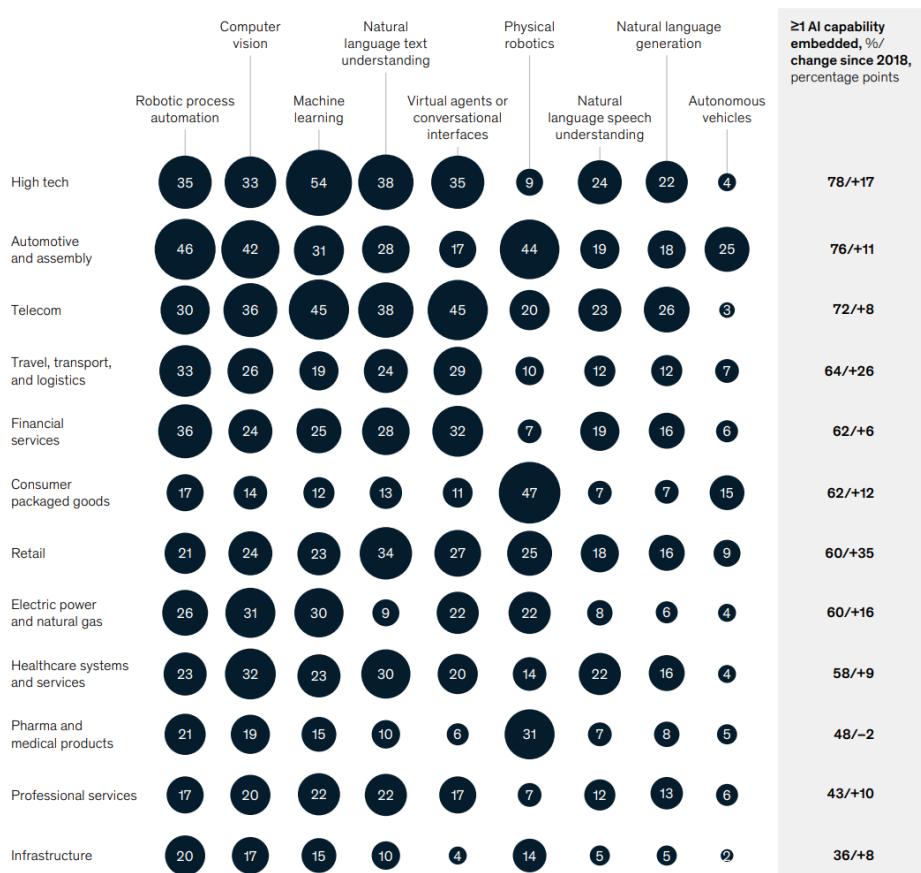


UNSW  
SYDNEY

# AI adoption in industry

High tech leads in AI adoption, and industries are generally using the AI capabilities most relevant to their value chains.

Organizations' AI capabilities,<sup>1</sup> % of respondents,<sup>2</sup> by industry



<sup>1</sup> Embedded in  $\geq 1$  product and/or business process for  $\geq 1$  function or business unit.

<sup>2</sup> Respondents who said "don't know" or "none of the above" are not shown. For high tech, n = 277; for automotive and assembly, n = 128; for telecom, n = 93; for travel, transport, and logistics, n = 83; for financial services, n = 396; for consumer packaged goods, n = 72; for retail, n = 94; for electric power and natural gas, n = 82; for healthcare systems and services, n = 78; for pharma and medical products, n = 96; for professional services, n = 331; and for infrastructure, n = 91.

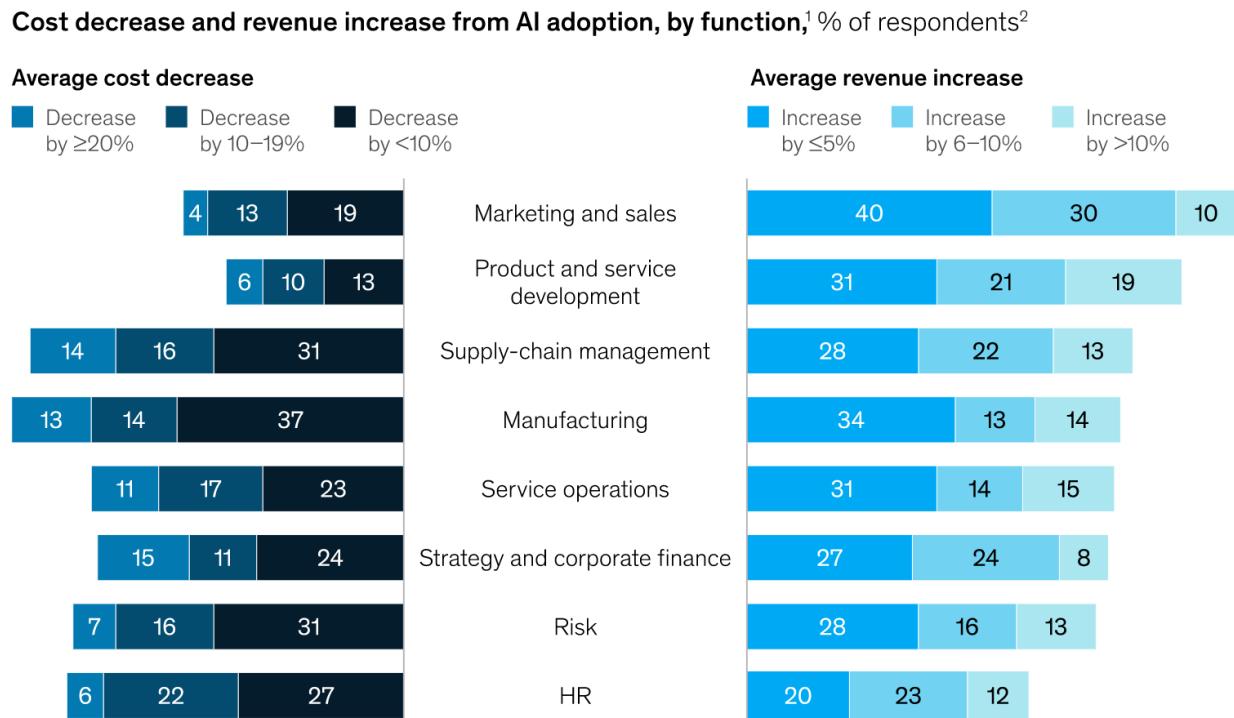
## AI global survey by McKinsey

<https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact>



# AI adoption leads to increased revenue and decreased cost

Revenue increases from adopting AI are reported most often in marketing and sales, and cost decreases most often in manufacturing.



<sup>1</sup>Marketing and sales includes the following use cases: customer-service analytics, customer segmentation, channel management, prediction of likelihood to buy, pricing and promotion, closed-loop marketing, marketing-budget allocation, churn reduction, and next product to buy. For product and service development: product-feature optimization, product-development-cycle optimization, creation of new AI-based enhancements, and creation of new AI-based products. For supply-chain management: logistics-network optimization, sales and parts forecasting, warehouse optimization, inventory and parts optimization, spend analytics, and sales and demand forecasting. For manufacturing: predictive maintenance and yield, energy, and throughput optimization. For service operations: service-operations optimization, contact-center automation, and predictive service and intervention. For strategy and corporate finance: capital allocation, treasury management, and M&A support. For risk: risk modeling/analytics, and fraud/debt analytics. For HR: performance management and organization-design, workforce-deployment, and talent-management optimization.

<sup>2</sup>Question asked only of respondents who said their companies adopted AI in given use case. Figures were calculated after removing respondents who said "don't know" or "not applicable; we are not tracking revenue related to AI"; respondents who said "no change" are not shown.

AI global survey by McKinsey

<https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact>



# Australia's AI Ethics Principles

Human, social  
and  
environmental  
wellbeing

Human-centred  
values

Fairness

Privacy  
protection and  
security

Reliability and  
safety

Transparency  
and explainability

Contestability

Accountability

Source: <https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles>



# Australia's AI Ethics Principles

**Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment.

**Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals.

**Fairness:** AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.

**Privacy protection and security:** AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.

**Reliability and safety:** AI systems should reliably operate in accordance with their intended purpose.

**Transparency and explainability:** There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.

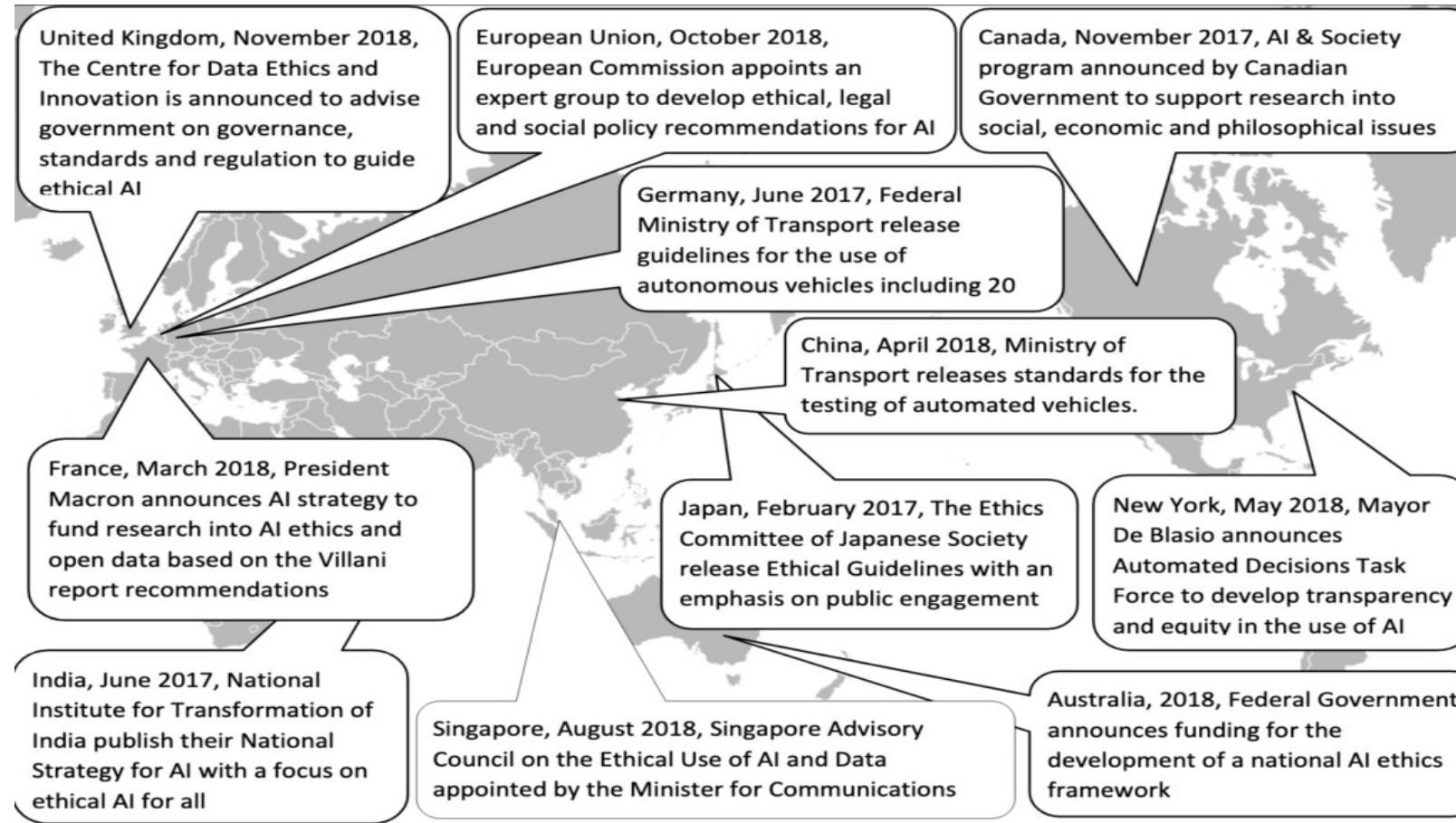
**Contestability:** When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.

**Accountability:** People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

Source: <https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles>



# A global snapshot of AI Ethics framework



Source: <https://www.csiro.au/-/media/D61/Reports/Artificial-Intelligence-ethics-framework.pdf>



# **Accountability, Privacy, Governance**



**UNSW**  
SYDNEY

# Reliability and Safety



Therac 25



UNSW  
SYDNEY

# Reliability? Case study: Robodebt

Centrelink debt recovery

Almost 200,000 robodebt cases to be wiped as Albanese government condemns 'shameful' scheme

Amanda Rishworth cites need for a welfare system that is 'not punitive' as she confirms reviews will be scrapped

Follow today's news, live

Get our free news app, morning email briefing or daily news podcast

Josh Butler

Tue 11 Oct 2022 22.31 AEDT



Social services minister Amanda Rishworth says the unlawful robodebt scheme was meant to save money but 'had a significant human cost'. Photograph: Lukas Coch/AAP

Nearly 200,000 Australians will have their robodebt reviews wiped as the federal government scraps investigations that were put on hold for three years under the unlawful welfare scheme.

About 124,000 people were told they were under review for social security payments they had received, while 73,000 were never informed they were being assessed for potential debts under the data-matching program.

## Robodebt

Not AI-based, but still an Automated Decision Making (ADM) system



UNSW  
SYDNEY

# Reliability? Case study: Robodebt

**A royal commission is set to review Robodebt. Its legacy is still hanging over Centrelink recipients today**

By the Specialist Reporting Team's Evan Young and Leonie Thorne  
Posted Sat 27 Aug 2022 at 5:04am, updated Sat 27 Aug 2022 at 5:21am



Emma Warren says she's still very affected by the Robodebt debacle today. (ABC News: Andrew Lobb)

Help keep family & friends informed by sharing this article

[abc.net.au/news/robodebt-royal-com...](https://www.abc.net.au/news/2022-08-27/robodebt-royal-commission-scars-linger/101375170)

 COPY LINK

 SHARE

It's been nearly six years, but Emma Warren still vividly remembers the moment she was told she owed the government \$8,000.

“Between 2015 and 2019, the scheme unlawfully claimed almost \$2 billion from more than 400,000 Australians. Some \$750 million was wrongly recovered from 381,000 people.”

<https://www.abc.net.au/news/2022-08-27/robodebt-royal-commission-scars-linger/101375170>



UNSW  
SYDNEY

# Reliability? Case study: Robodebt

**A royal commission is set to review Robodebt. Its legacy is still hanging over Centrelink recipients today**

By the Specialist Reporting Team's Evan Young and Leonie Thorne  
Posted Sat 27 Aug 2022 at 5:04am, updated Sat 27 Aug 2022 at 5:21am



Emma Warren says she's still very affected by the Robodebt debacle today. (ABC News: Andrew Lobb)

Help keep family & friends informed by sharing this article

[abc.net.au/news/robodebt-royal-com...](https://abc.net.au/news/robodebt-royal-com...)

 COPY LINK

 SHARE

It's been nearly six years, but Emma Warren still vividly remembers the moment she was told she owed the government \$8,000.

On 11 Jun 2021 the Federal Court approved a A\$1.872 billion settlement incorporating repayment of A\$751 million, wiping of all remaining debts, and the legal costs running to A\$8.4 million

<https://www.abc.net.au/news/2021-06-11/robodebt-condemned-by-federal-court-judge-as-shameful-chapter/100207674>



UNSW  
SYDNEY

# Accountability

Four barriers:

1. The problem of many hands
2. The problem of bugs
3. Blaming the computer
4. Software ownership without liability

Nissenbaum, H., 2020. Computing and accountability. In *The Ethics of Information Technologies* (pp. 351-358). Routledge.

Nissenbaum H. Accountability in a computerized society. Science and engineering ethics. 1996 Mar;2(1):25-42.

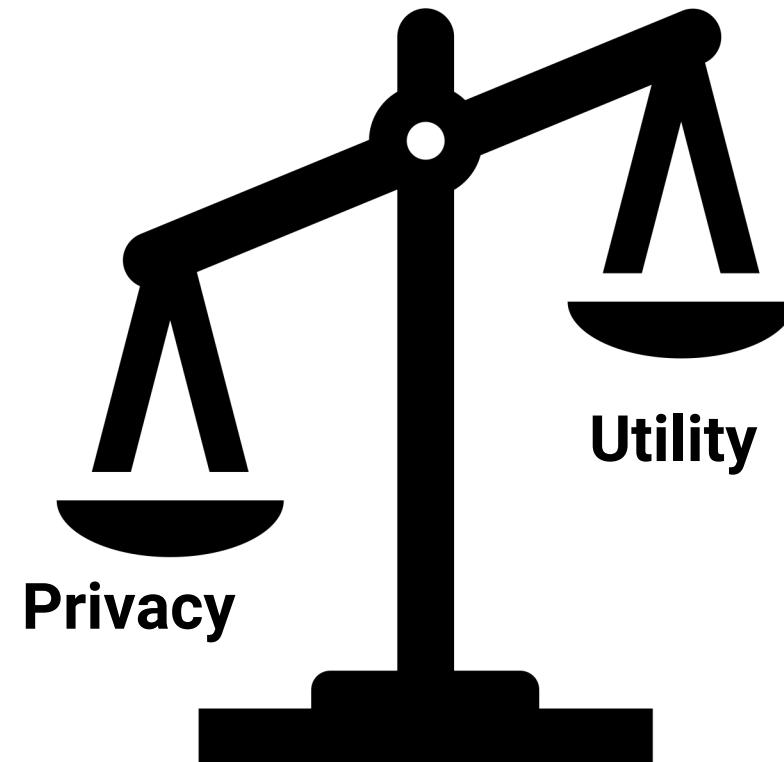


# Data: the core of AI

*“The recent advances in key AI capabilities such as deep learning have been made possible by vast troves of data. This data has to be collected and used, which means issues related to AI are closely intertwined with those that relate to privacy and data.”* - CSIRO & Data61, Innovation, Department of Industry, and Australian Government Science



# Privacy vs. Utility Tradeoff



# Measuring Privacy

- Secrecy: it concerns information that others may gather about us
  - the probability of a data item being accessed
  - the change in knowledge of an adversary/third party upon seeing the data
- Anonymity: it addresses how much in the public gaze we are
  - the privacy leakage is measured in terms of the size of the blurring accompanying the release of data
- Solitude: it measures the degree to which others have physical access to us

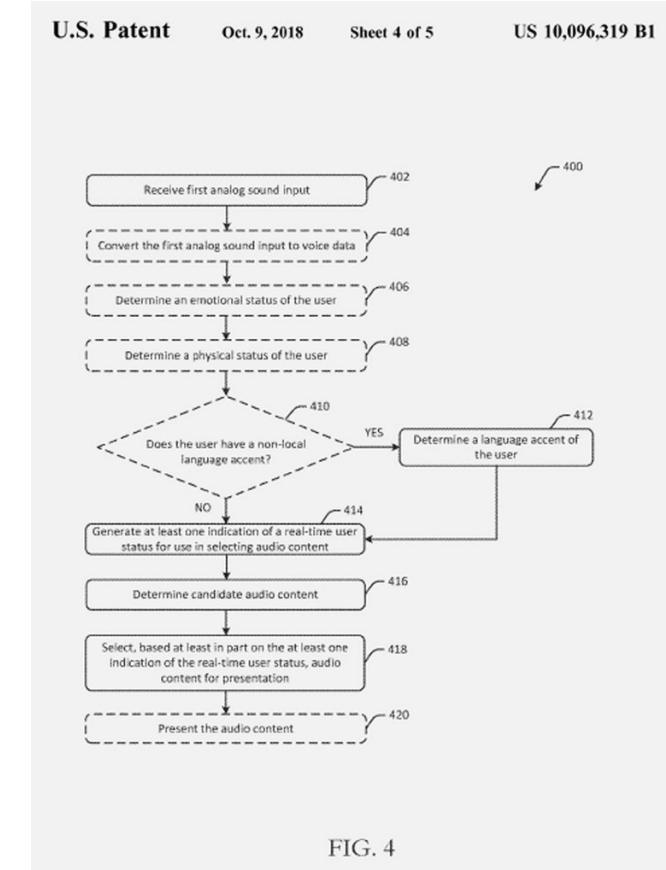
(Xintao Wu, University of Arkansas)



# Case study: AMAZON'S ACCENT RECOGNITION TECHNOLOGY

AT THE BEGINNING of October 2018, Amazon was quietly issued a patent that would allow its virtual assistant Alexa to decipher a user's physical characteristics and emotional state based on their voice. Characteristics, or "voice features," like language accent, ethnic origin, emotion, gender, age, and background noise would be immediately extracted and tagged to the user's data file to help deliver more targeted advertising.

The algorithm would also consider a customer's physical location — based on their IP address, primary shipping address, and browser settings — to help determine their accent.



The illustration from Document:  
United States Patent and  
Trademark Office



UNSW  
SYDNEY

# Discuss

- What are the potential ethical concerns in relation to data privacy?
- Are there any concerns with regards to the intended use of this data and analysis?
- Will AI models be limited if users don't want to provide data?
- Are there any other potential issues/concerns regarding FATE (Fairness, Accountability, Transparency, and Ethics)?



# Consent and the Privacy Act

- Four key terms
  - The individual is adequately informed before giving consent.
  - The individual gives consent voluntarily.
  - The consent is current and specific.
  - The individual has the capacity to understand and communicate their consent.
- The ‘right to be **forgotten**’



# Case study: Cambridge Analytica and public trust

Through a Facebook app, a Cambridge University researcher was able to gain access to the personal information of not only users who agreed to take the survey, but also the people in those users' Facebook social networks. In this way, the app harvested data from millions of Facebook users. Various reports indicate that these data were then used to develop targeted advertising for various political campaigns run by Cambridge Analytica.



# Case study: Cambridge Analytica and public trust

- Are there potential ethical concerns in relation to data privacy?
- Are there any concerns with regards to the intended use of this data and analysis for advertisement?



# Case study: Cambridge Analytica and public trust

- Lesson we learned
  - Cost of inadequate data protection policies
  - Not sufficient to merely follow the letter of the law
- Code of practices
  - Australian Privacy Act
  - Consent process: current, specific, transparent
  - Regular review of data collection and usage policies
  - Balance between privacy protection and technology innovation

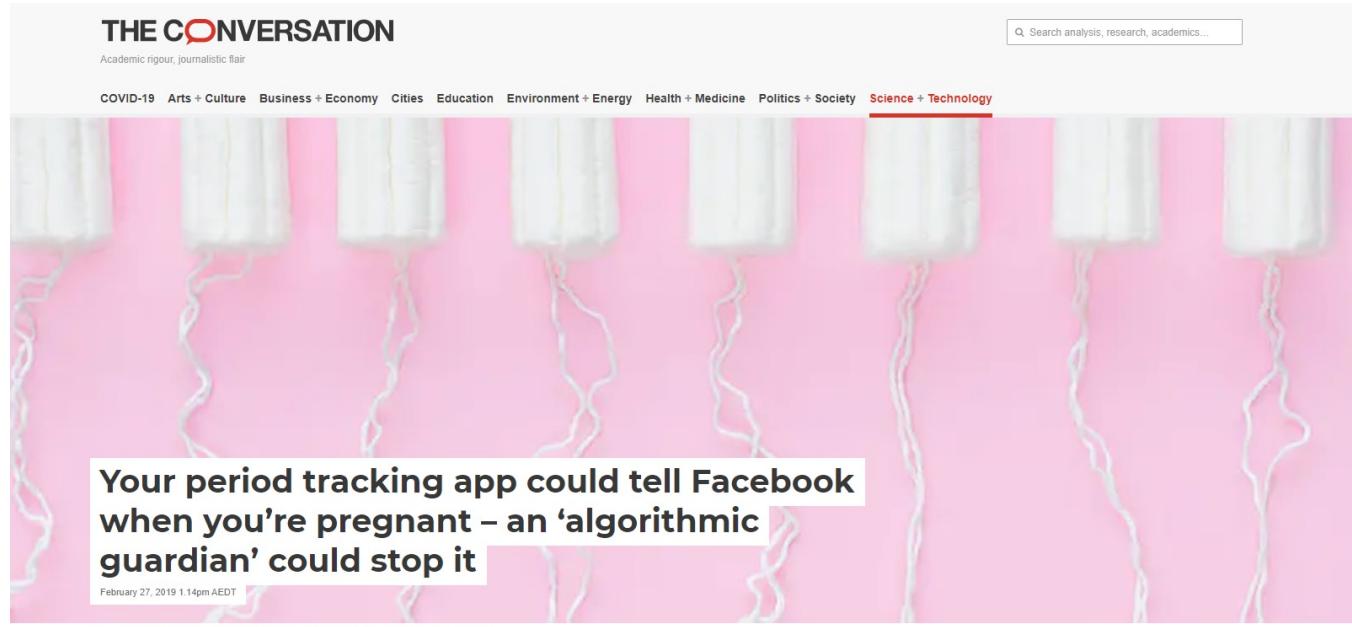


# Case Study: Mood and Perception Manipulation

A controversial research study (Kramer, Guillory, and Hancock [2014](#)) used Facebook's platform to demonstrate that users' moods can be manipulated by filtering their feeds (comments, videos, pictures and web links posted by their Facebook friends). The social media company altered the news feeds (the main page users land on for a stream of updates from friends) of nearly 700,000 users. Feeds were changed to reflect more "positive" or "negative" content, to determine if seeing more sad messages makes a person sadder. The study shows that reducing exposure to feeds with positive content led to the user posting fewer positive posts, and the same pattern occurred for negative content.



# Case study: Period tracking app



The way it is now, we have almost no way to know how our data are being shared and used. Shutterstock

Email

Twitter

Facebook

LinkedIn

Print

15

47

Most of us know tech platforms such as Facebook and Google track, store and make money from our data. But there are constantly new revelations about just how much of our privacy has been chipped away.

The latest comes from the Wall Street Journal, which [dropped a bombshell](#) on Friday when its testing revealed many popular smartphone apps have been sending personal data to Facebook. That reportedly includes data from heart rate monitoring and period tracking apps:

Flo Health Inc's Flo Period & Ovulation Tracker, which claims 25 million active users, told Facebook when a user was having her period or informed the app of an intention to get pregnant, the tests showed.

When we use technologies that track our data, we enter a system governed by algorithms. And the more information we hand over, the more we become entrained with algorithmic systems we don't control.

Authors

 Flora D. Salim  
Associate Professor in Computer Science, RMIT University

 Salil S. Kanhere  
Professor, UNSW

 Seng W. Loke  
Professor In Computer Science, Deakin University

Disclosure statement

The authors do not work for, consult, own shares in or receive funding from any company or organisation that would benefit from this article, and have disclosed no relevant affiliations beyond their academic appointment.

Partners

<https://theconversation.com/your-period-tracking-app-could-tell-facebook-when-youre-pregnant-an-algorithmic-guardian-could-stop-it-111815>



UNSW

# Data Security and Governance

- Establishes policies around data that are used for operational purposes in an organization.
- It defines actions on data in particular situations, i.e. who is going to take action and what methods will be used for this action.
- A good data governance framework ensures:
  - data quality
  - data availability
  - data protection.



# Data Security and Governance

- Avoid the \$\$\$ cost of data breach
- And the massive cost of the loss of reputation



UNSW

# Case Study: Equifax data breach

In 2017 Equifax, a US-based credit reporting agency, experienced a data breach affecting at least 145.5 million individuals, with various degrees of sensitive personal information compromised. In addition, due to the huge number of people affected, it took several weeks to identify the individuals and notify the public that the breach had occurred. The cost of the breach was estimated to be in the realm of US\$275 million.



# Transparency



**UNSW**  
SYDNEY  
**UNSW**  
SYDNEY

# Transparency in AI

Transparency is one of seven key requirements for the realisation of ‘trustworthy AI’ (EU Commission’s High-Level Expert Group on AI (AI HLEG) in April 2019 )

“Transparency” is the single most common, and one of the key five principles emphasised in the vast number – a recent study counted 84 – of ethical guidelines addressing AI on a global level (Jobin et al., 2019).

Larsson, S. & Heintz, F. (2020). Transparency in artificial intelligence. Internet Policy Review, 9(2).  
<https://doi.org/10.14763/2020.2.1469>; <https://policyreview.info/concepts/transparency-artificial-intelligence>

# Black box vs white box algorithms



vs.



Norbert Wiener, 1948, Cybernetics: or Control and Communication in the Animal and the Machine



UNSW

# The need for Transparency and Explainable AI (XAI)

- The problems of accountability as computing technologies becoming more complex and less intelligible (Helen Nissenbaum).
- The opacity in Machine Learning Systems (Jean Burrell, 2016) due to :
  - Trade secrets
  - Limited people with the knowledge of programming languages and ML
  - The complexity and high dimensionality of data for decision making no longer match human-scale reasoning
- Institutional transparency, public values, regulations
  - Customer's rights for explanations (GDPR Article 15(1))
  - Requirement for human in the loop (GDPR Article 22)
  - Requirement for algorithmic auditing (US Algorithmic Accountability act)

Source: Jake Goldenfein, 'Algorithmic Transparency and Decision-Making Accountability: Thoughts for buying machine learning algorithms' in Office of the Victorian Information Commissioner (ed), Closer to the Machine: Technical, Social, and Legal aspects of AI (2019), Available at SSRN: <https://ssrn.com/abstract=3445873>, <https://ovic.vic.gov.au/wp-content/uploads/2019/08/closer-to-the-machine-web.pdf> p.45-65

# Transparency and Explainability

More on Week 10



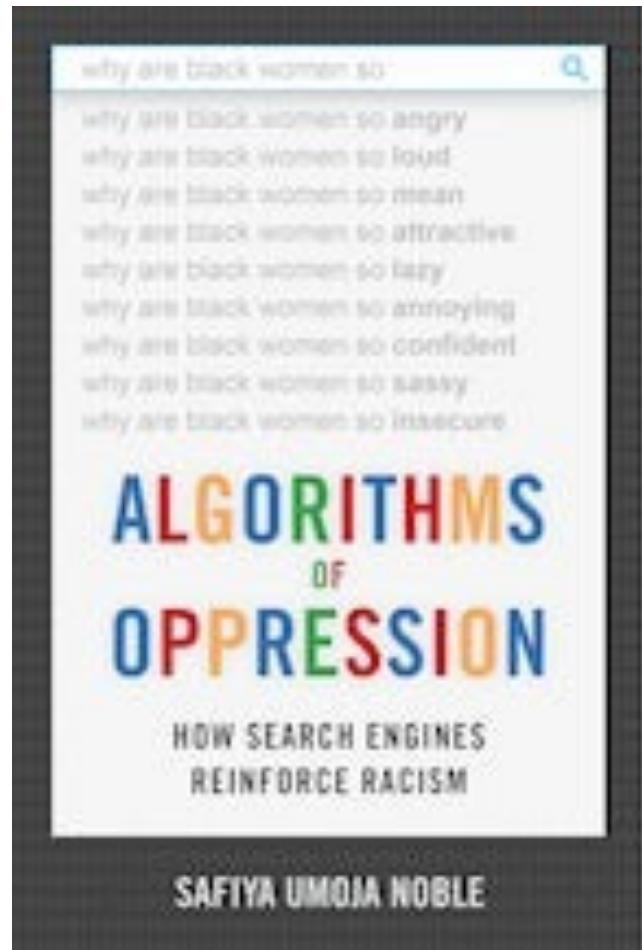
UNSW  
SYDNEY

# Fairness



**UNSW**  
SYDNEY  
**UNSW**  
SYDNEY

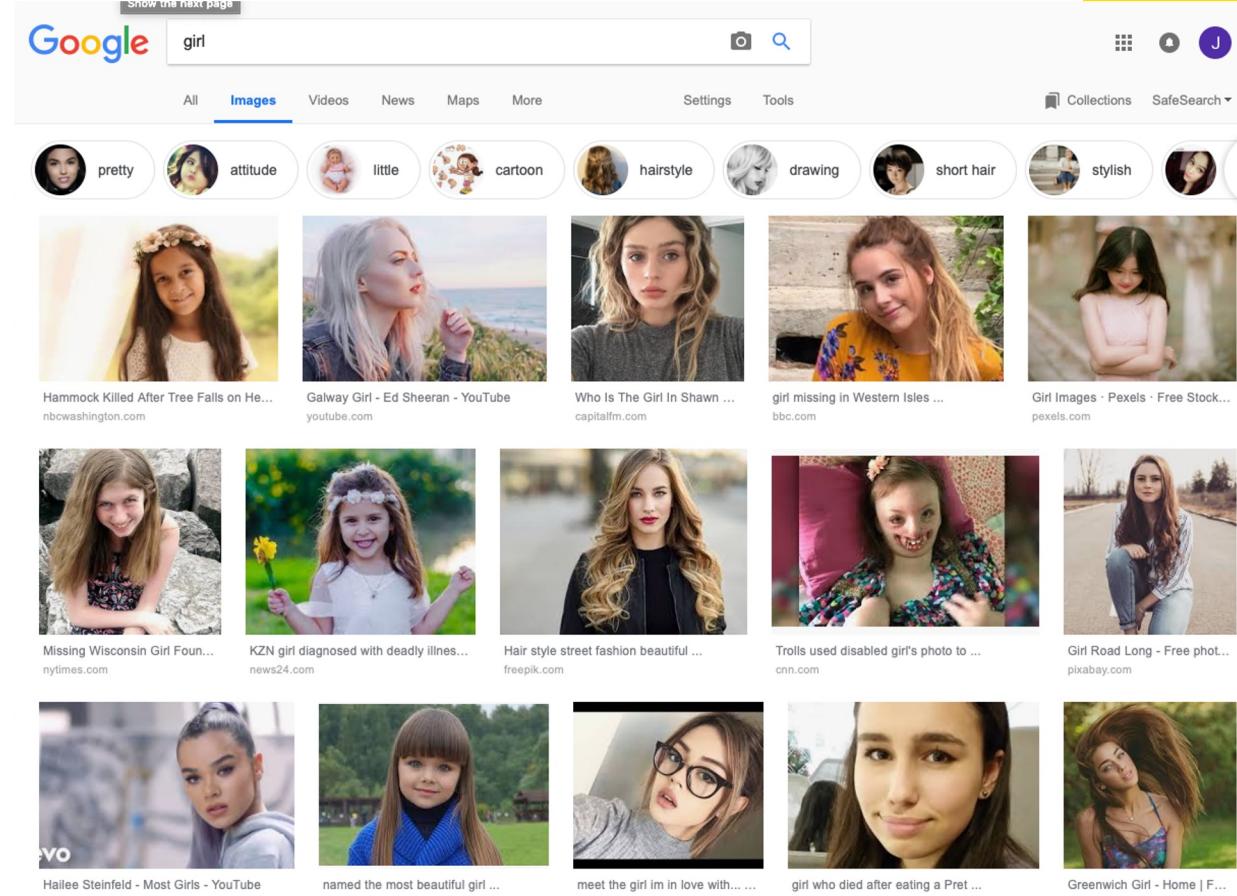
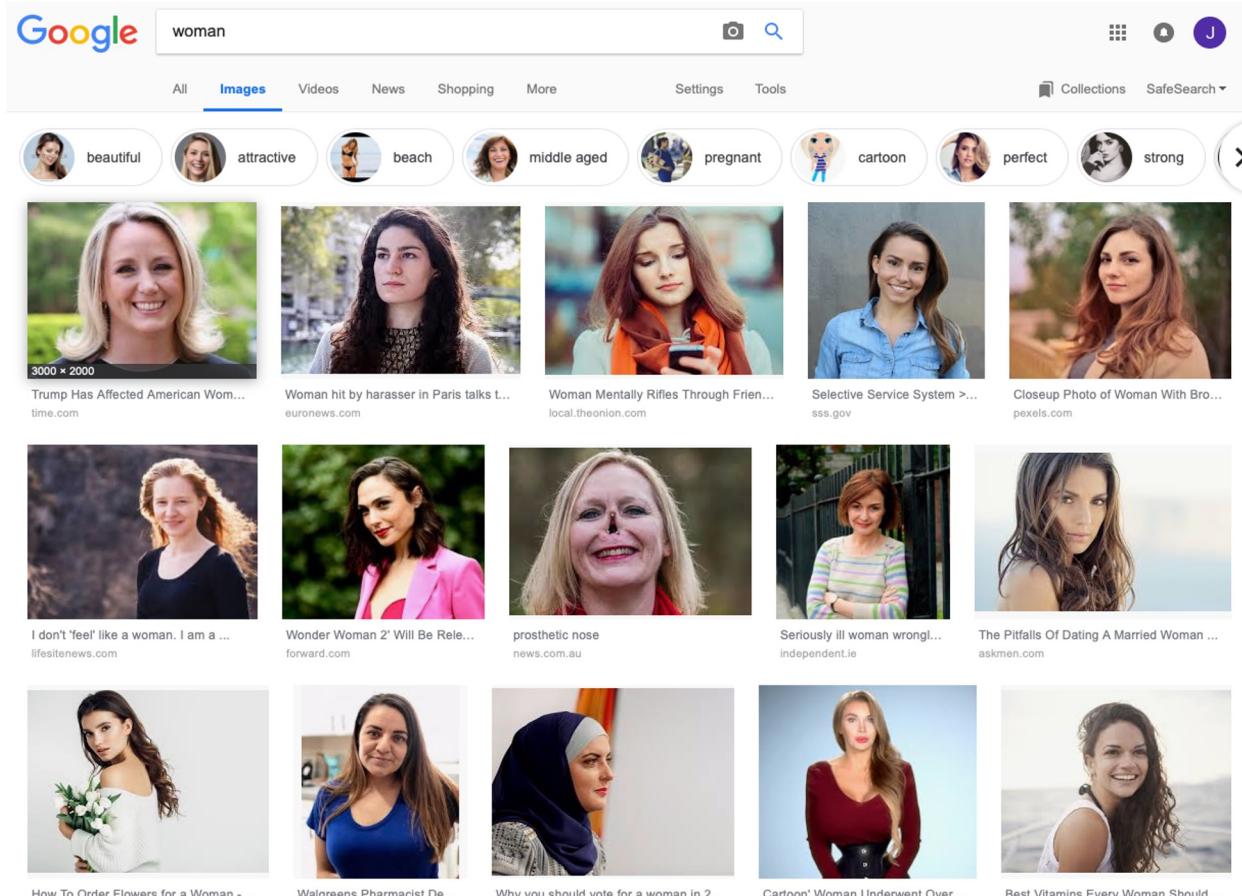
# Algorithms of Oppression



Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.



# Search Engine Bias



Source from <https://scroll.in/article/921305/google-search-results-reflect-its-algorithms-bias-against-women-and-people-of-colour>

# Search Engine result on CEO (3 years ago)

← → C google.com/search?q=CEO&liz=1C5CHFA\_enAU874AU874&source=lnms&tbo=isch&sa=X&ved=2ahUKEwjAkZGco6\_mAhVTILcAHV8oAyBQ\_AUoAXoECA8QAw&biw=1501&bih=764

Paused M

Google CEO

All Images News Maps Videos More Settings Tools Collections SafeSearch ▾

business google cartoon snapchat microsoft apple woman desk amazon uber

The image shows a Google search results page for the query "CEO". The top navigation bar includes links for All, Images (which is selected), News, Maps, Videos, More, Settings, Tools, Collections, and SafeSearch. Below the search bar is a row of circular icons representing different search terms: business, google, cartoon, snapchat, microsoft, apple, woman, desk, amazon, and uber. The main content area displays ten search results, each consisting of a thumbnail image and a brief description. The results include a group photo of CEOs, a woman in a business suit, a man in a suit, a man in a tuxedo at an awards ceremony, a woman in a suit, a man in a suit, a man in a suit, a man in a suit, a woman in a suit, and a man in a suit.

Chief executive officer - Wikipedia  
en.wikipedia.org

What do CEOs do? A CEO Job Description ...  
steverobbins.com

ABB Turns to New CEO Rosengren to Drive ...  
industryweek.com

Marriott CEO Arne Sorenson Is The 2...  
chiefexecutive.net

Mark Nielsen named Australian CEO of ...  
talentinternational.com.au

How to use 'CEO magic' when trying to ...  
europeanceo.com

Why Aviva's CEO Mark Wilson has fallen ...  
independent.co.uk

BHP | Mike Henry to become BHP Chief ...  
bhp.com

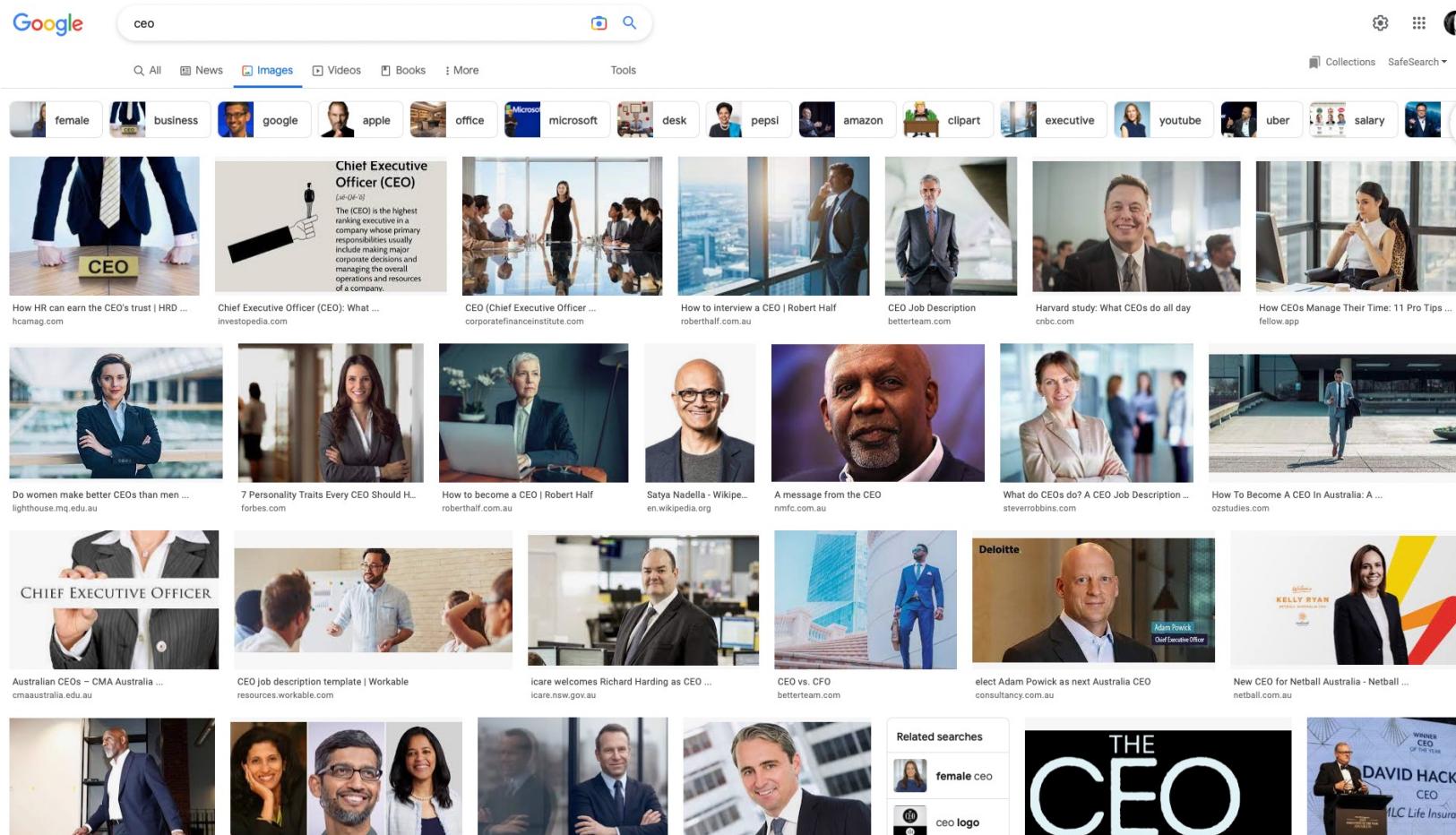
PwC's 22nd CEO Survey - CEO's curbed ...  
pwc.com

You are the CEO of Your Life | Personal ...  
personalexcellence.co



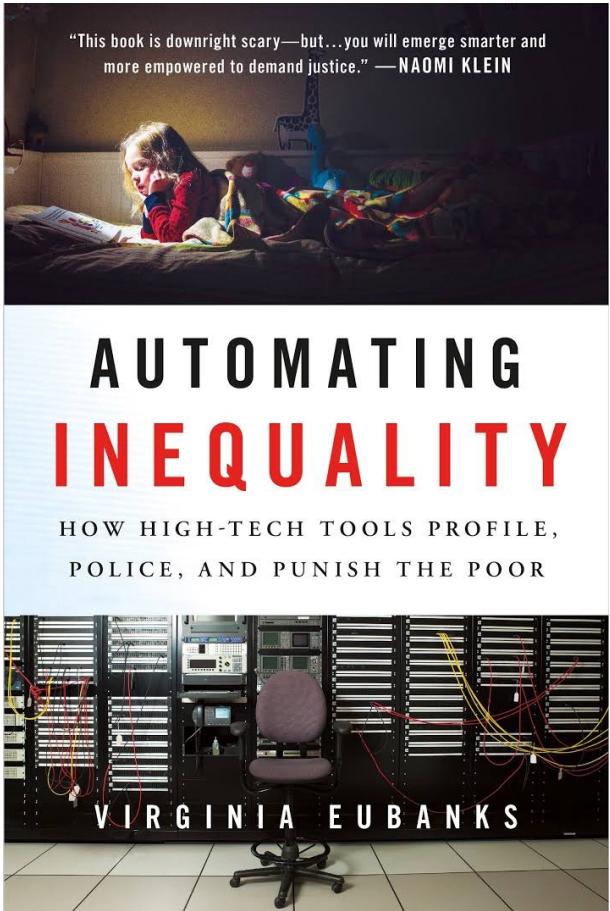
UNSW  
SYDNEY

# Search Engine result on CEO (today)



UNSW  
SYDNEY

# Virginia Eubanks



Virginia Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

An excerpt:  
<https://www.scientificamerican.com/article/algorithms-designed-to-fight-poverty-can-actually-make-it-worse/>



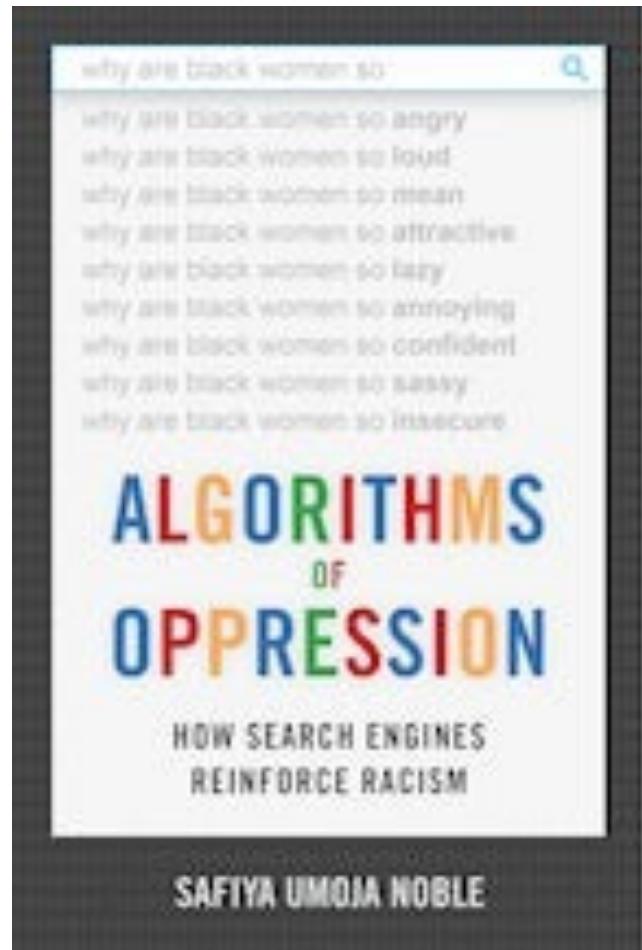
UNSW  
SYDNEY

# Fairness



**UNSW**  
SYDNEY  
**UNSW**  
SYDNEY

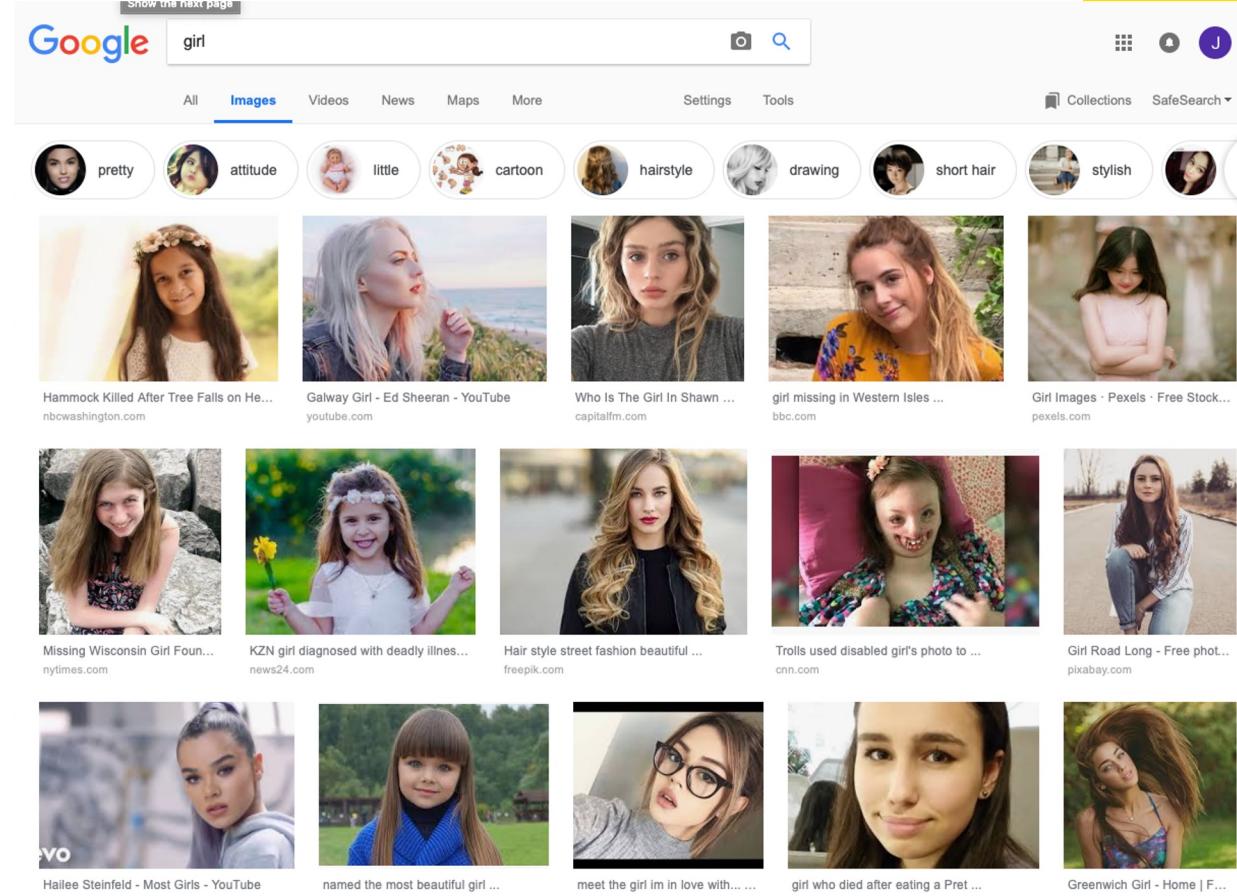
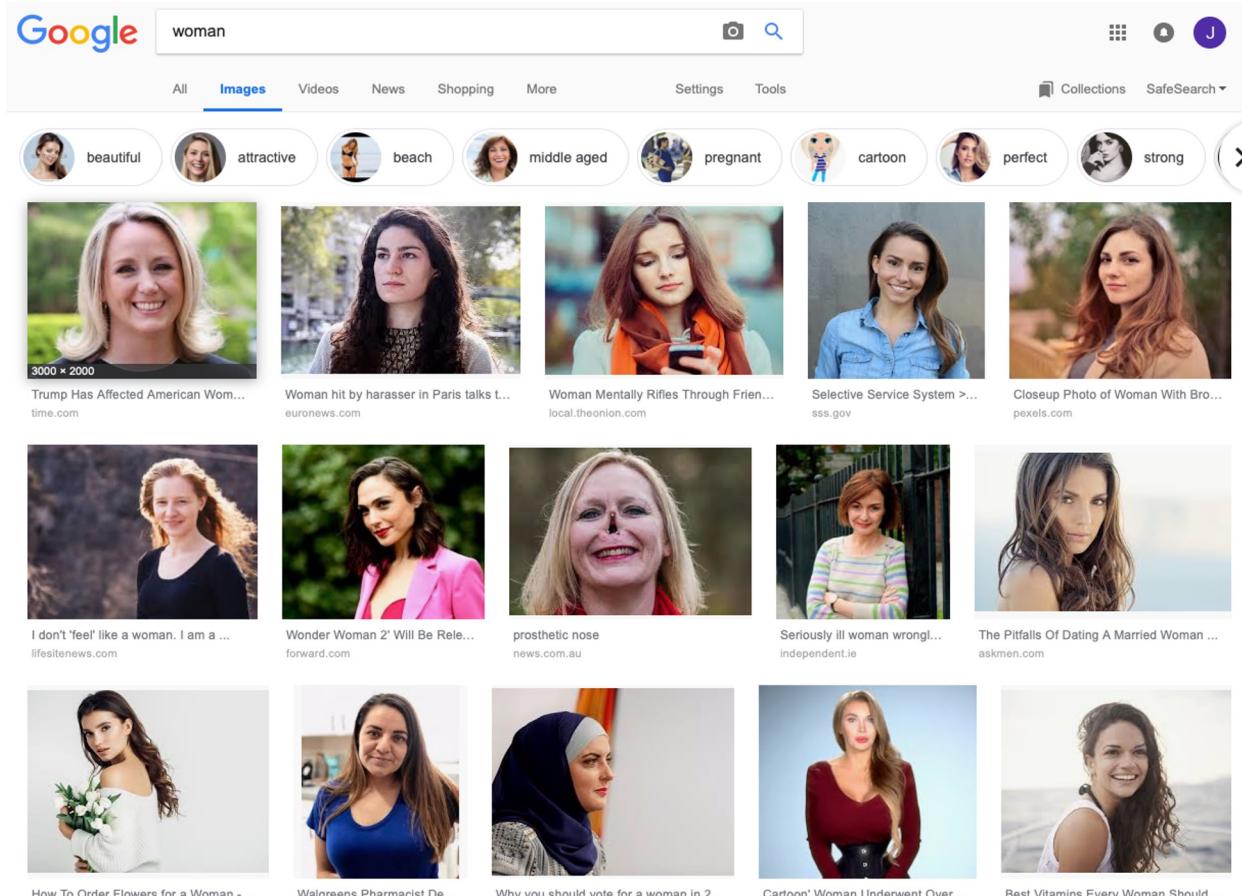
# Algorithms of Oppression



Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.



# Search Engine Bias



Source from <https://scroll.in/article/921305/google-search-results-reflect-its-algorithms-bias-against-women-and-people-of-colour>

# Search Engine result on CEO (3 years ago)

← → C google.com/search?q=CEO&liz=1C5CHFA\_enAU874AU874&source=lnms&tbo=isch&sa=X&ved=2ahUKEwjAkZGco6\_mAhVTILcAHV8oAyBQ\_AUoAXoECA8QAw&biw=1501&bih=764

Paused M

Google CEO

All Images News Maps Videos More Settings Tools Collections SafeSearch ▾

business google cartoon snapchat microsoft apple woman desk amazon uber

The image shows a Google search results page for the query "CEO". The top navigation bar includes links for All, Images (which is selected), News, Maps, Videos, More, Settings, Tools, Collections, and SafeSearch. Below the search bar is a row of circular icons representing different search terms: business, google, cartoon, snapchat, microsoft, apple, woman, desk, amazon, and uber. The main content area displays ten image search results, each with a thumbnail, a caption, and a link to the source website. The results include a group photo of CEOs, individual portraits of CEOs like Arne Sorenson and Mark Nielsen, and articles about CEOs like Mike Henry and Mark Wilson.

Chief executive officer - Wikipedia  
en.wikipedia.org

What do CEOs do? A CEO Job Description ...  
steverobbins.com

ABB Turns to New CEO Rosengren to Drive ...  
industryweek.com

Marriott CEO Arne Sorenson Is The 2...  
chiefexecutive.net

Mark Nielsen named Australian CEO of ...  
talentinternational.com.au

How to use 'CEO magic' when trying to ...  
europeanceo.com

Why Aviva's CEO Mark Wilson has fallen ...  
independent.co.uk

BHP | Mike Henry to become BHP Chief ...  
bhp.com

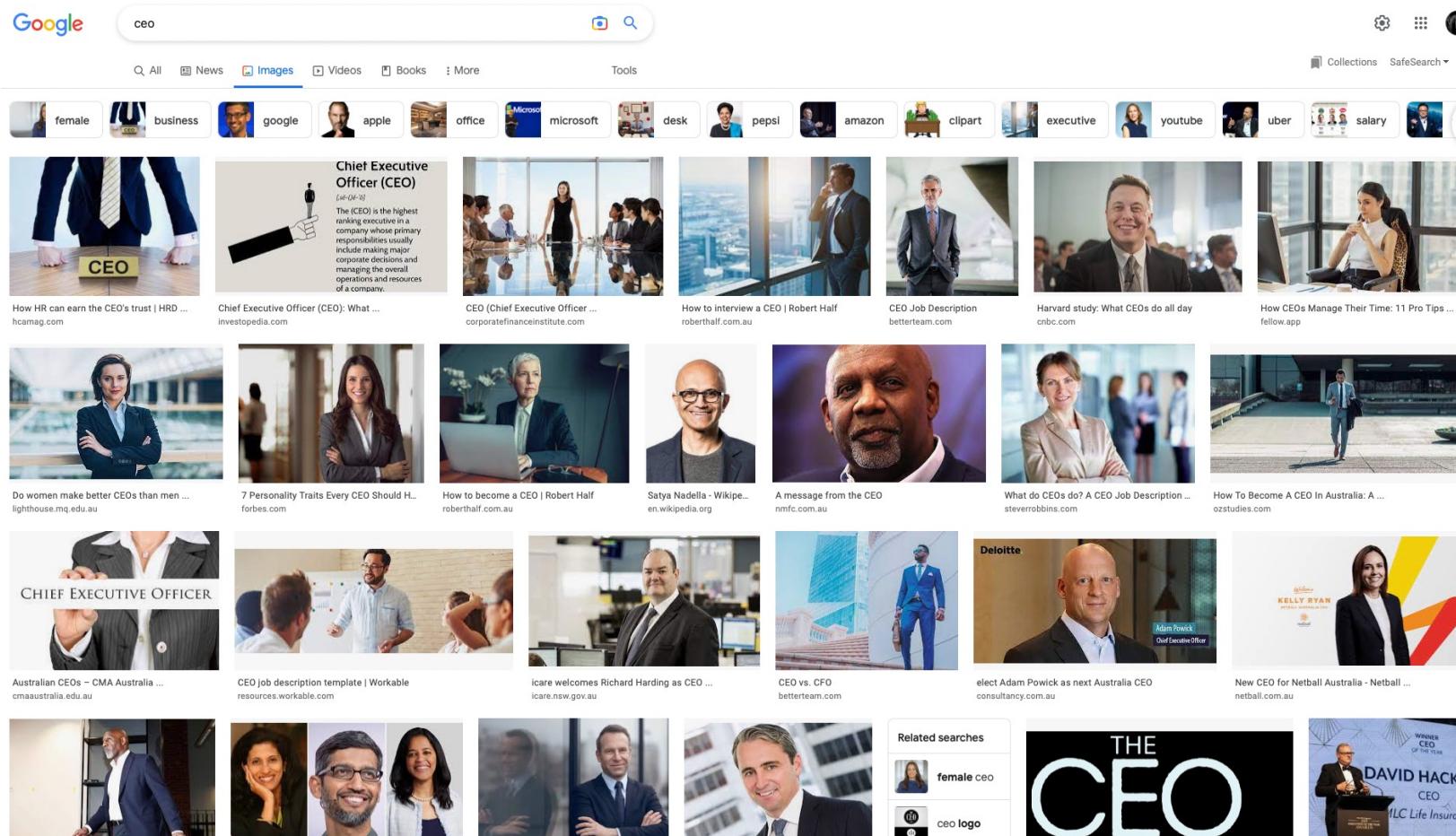
PwC's 22nd CEO Survey - CEO's curbed ...  
pwc.com

You are the CEO of Your Life | Personal ...  
personalexcellence.co



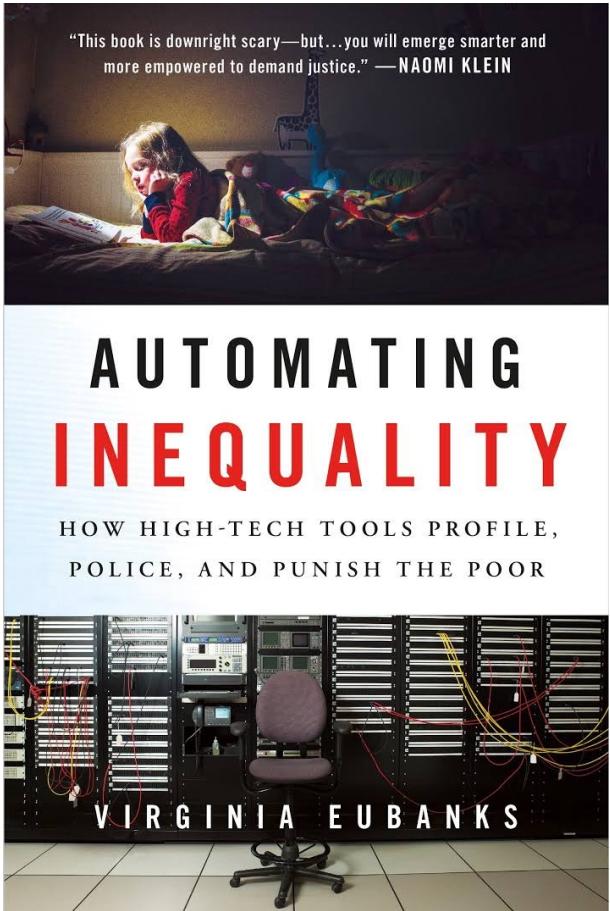
UNSW  
SYDNEY

# Search Engine result on CEO (today)



UNSW  
SYDNEY

# Virginia Eubanks



Virginia Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

An excerpt:  
<https://www.scientificamerican.com/article/algorithms-designed-to-fight-poverty-can-actually-make-it-worse/>



UNSW  
SYDNEY

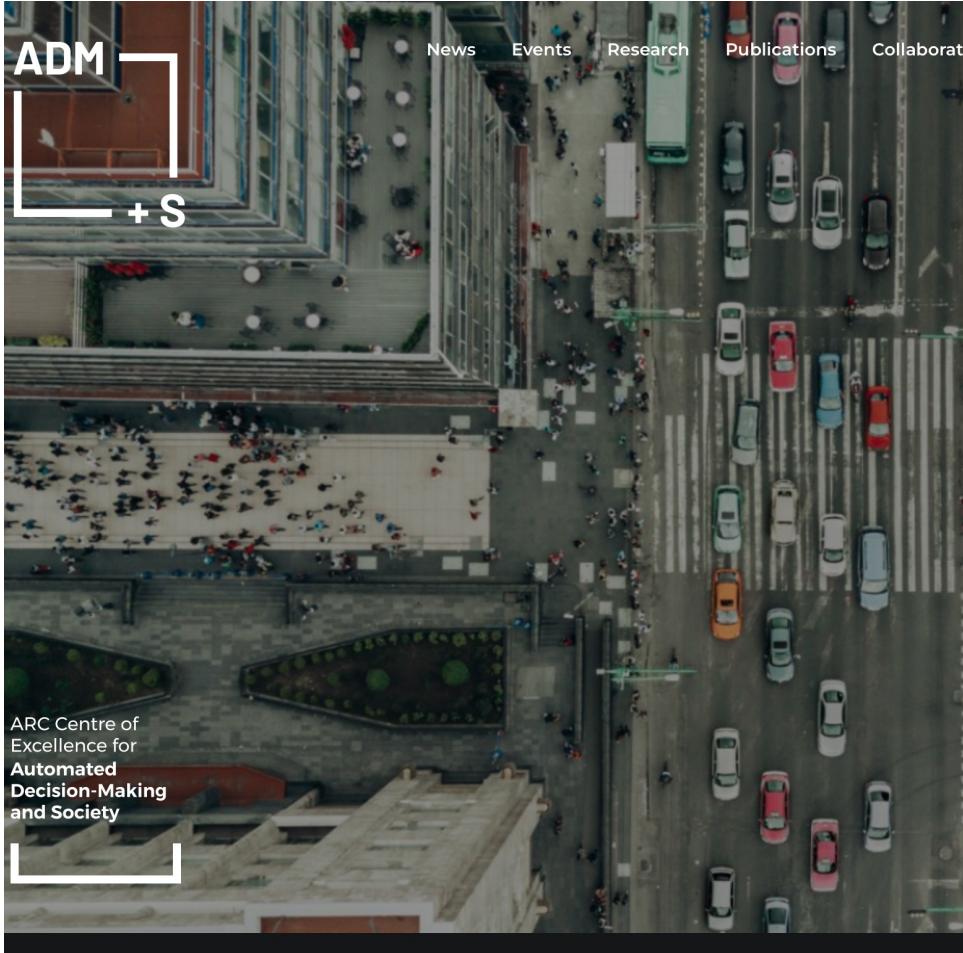
# Bias and Fairness

More on Week 7



UNSW  
SYDNEY

# ADM+S Centre



The website features a large, high-angle photograph of a modern urban building complex. In the top left corner of the image, there is a white graphic element consisting of a vertical bar with the letters 'ADM' at the top and a horizontal bar with the letters '+ S' at the bottom, forming a stylized 'L' shape.

News   Events   Research   Publications   Collaborate   About Us    Search

CREATING KNOWLEDGE AND STRATEGIES FOR RESPONSIBLE, ETHICAL AND INCLUSIVE AUTOMATED DECISION-MAKING.

LATEST NEWS

Future Automated Mobilities: Towards hope, justice, and care  
Participate in an interdisciplinary symposium on October 20-21 to envisage safe and trusted automated mobility futures with leading experts.

< || >

Sign up for regular updates →

<https://www.admscentre.org.au>

# Further readings

Friedman B, Kahn PH, Borning A, Hultgren A. Value sensitive design and information systems. In Early engagement and new technologies: Opening up the laboratory 2013 (pp. 55-95). Springer, Dordrecht.

