Data Ethics Lecture 3

recap **Promises**

preview **more promises**

Ali Alkhatib
@_alialkhatib || hi@al2.in

March 24, 2022

Roadmap for today

- Admin?
- Recap Promises
- Preview other promises (tedious/dangerous & social good/token stakeholders)

ADMIN STUFF?

- reading time/reflection

ADMIN STUFF?

- reading time/reflection
- let’s do that again →

check the zoom chat for a link

Promises

ROADMAP

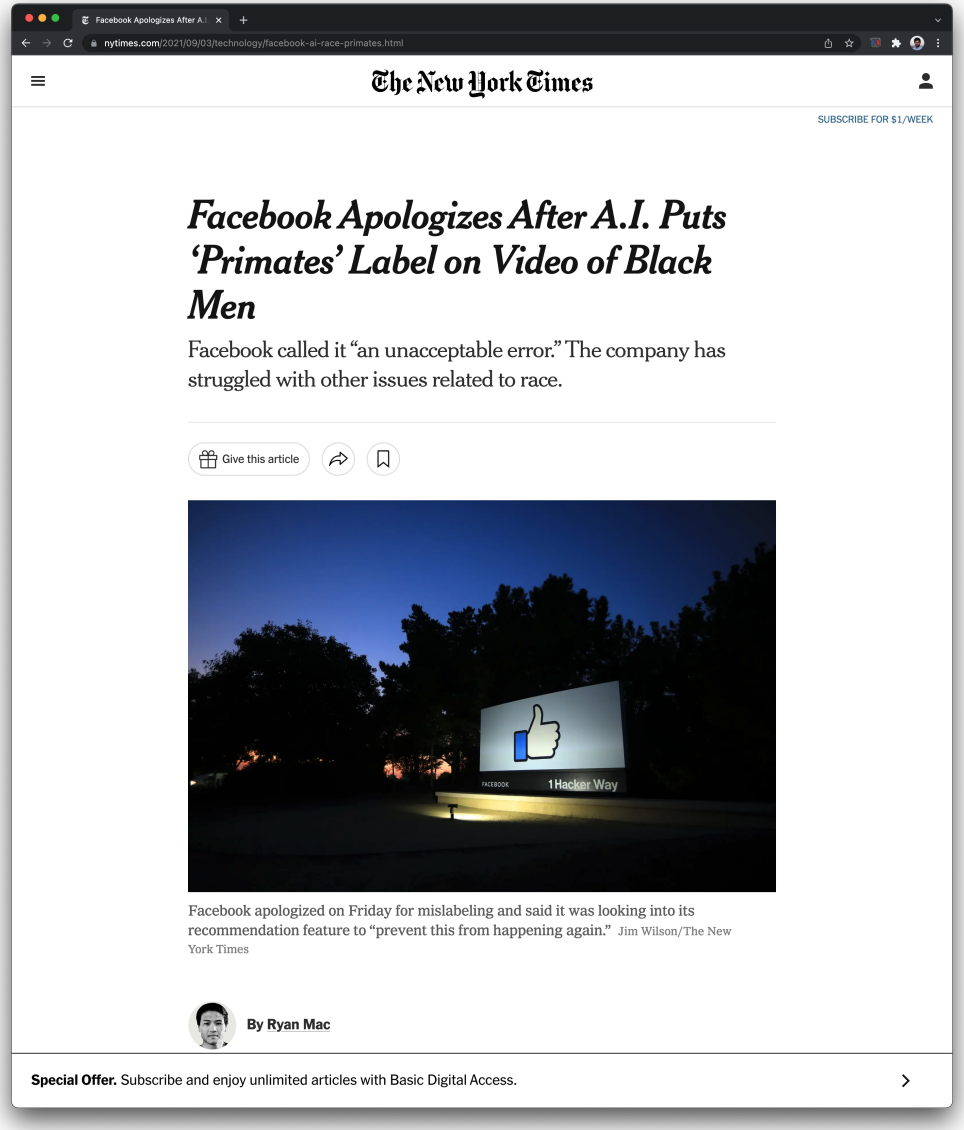
- AI can “understand” the world better than we can
- AI can make “fairer” decisions than we can

Understanding

AI WILL "UNDERSTAND" BETTER THAN WE CAN

AI WILL UNDERSTAND MORE

→ labeling



Facebook Apologizes After A.I.

nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html

The New York Times

SUBSCRIBE FOR $1/WEEK

Facebook Apologizes After A.I. Puts 'Primates' Label on Video of Black Men

Facebook called it "an unacceptable error." The company has struggled with other issues related to race.

Give this article

Facebook apologized on Friday for mislabeling and said it was looking into its recommendation feature to "prevent this from happening again." Jim Wilson/The New York Times

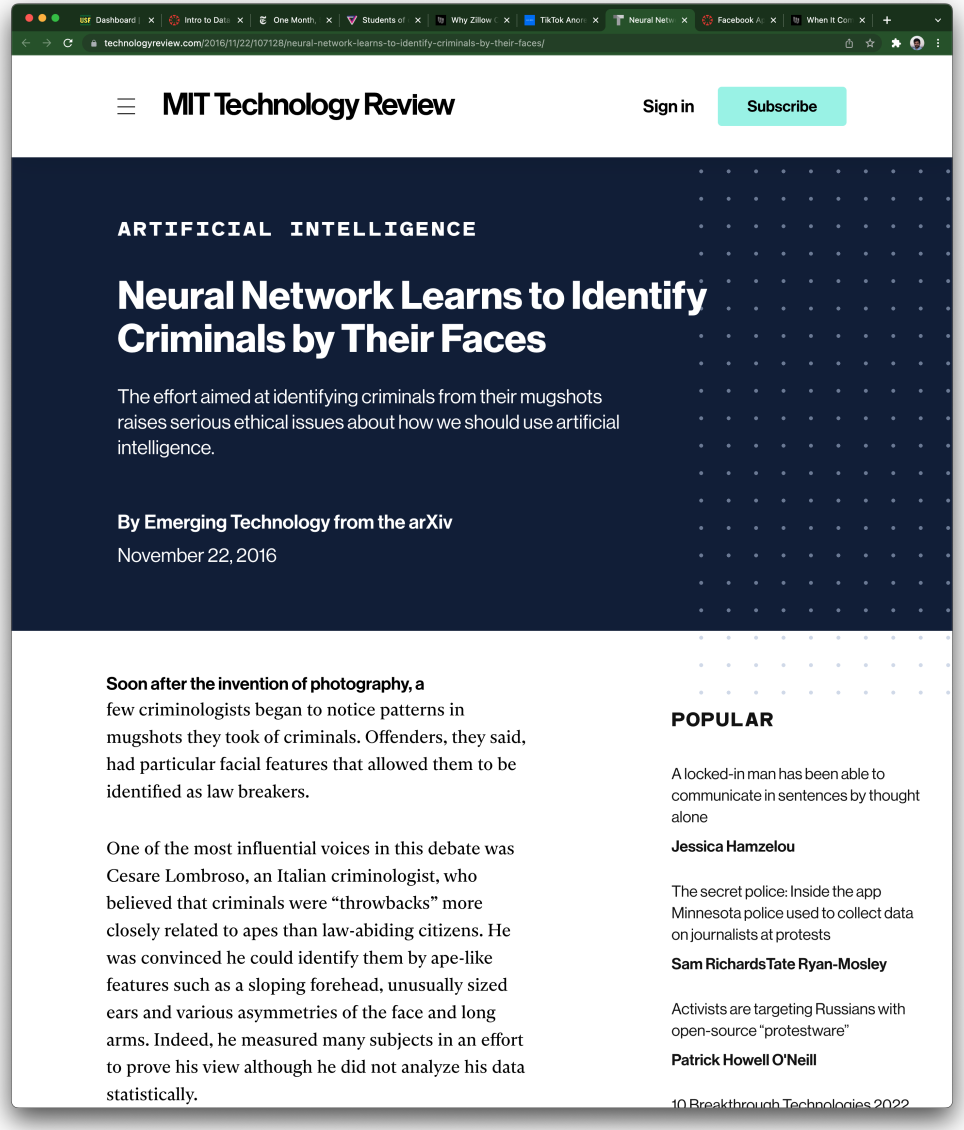
By Ryan Mac

AI WILL UNDERSTAND MORE

→ labeling

AI WILL UNDERSTAND MORE

→ labeling



technologyreview.com/2016/11/22/107128/neural-network-learns-to-identify-criminals-by-their-faces/

MIT Technology Review

Sign in Subscribe

ARTIFICIAL INTELLIGENCE

Neural Network Learns to Identify Criminals by Their Faces

The effort aimed at identifying criminals from their mugshots raises serious ethical issues about how we should use artificial intelligence.

By Emerging Technology from the arXiv

November 22, 2016

Soon after the invention of photography, a few criminologists began to notice patterns in mugshots they took of criminals. Offenders, they said, had particular facial features that allowed them to be identified as law breakers.

One of the most influential voices in this debate was Cesare Lombroso, an Italian criminologist, who believed that criminals were "throwbacks" more closely related to apes than law-abiding citizens. He was convinced he could identify them by ape-like features such as a sloping forehead, unusually sized ears and various asymmetries of the face and long arms. Indeed, he measured many subjects in an effort to prove his view although he did not analyze his data statistically.

POPULAR

A locked-in man has been able to communicate in sentences by thought alone

Jessica Hamzelou

The secret police: Inside the app Minnesota police used to collect data on journalists at protests

Sam RichardsTate Ryan-Mosley

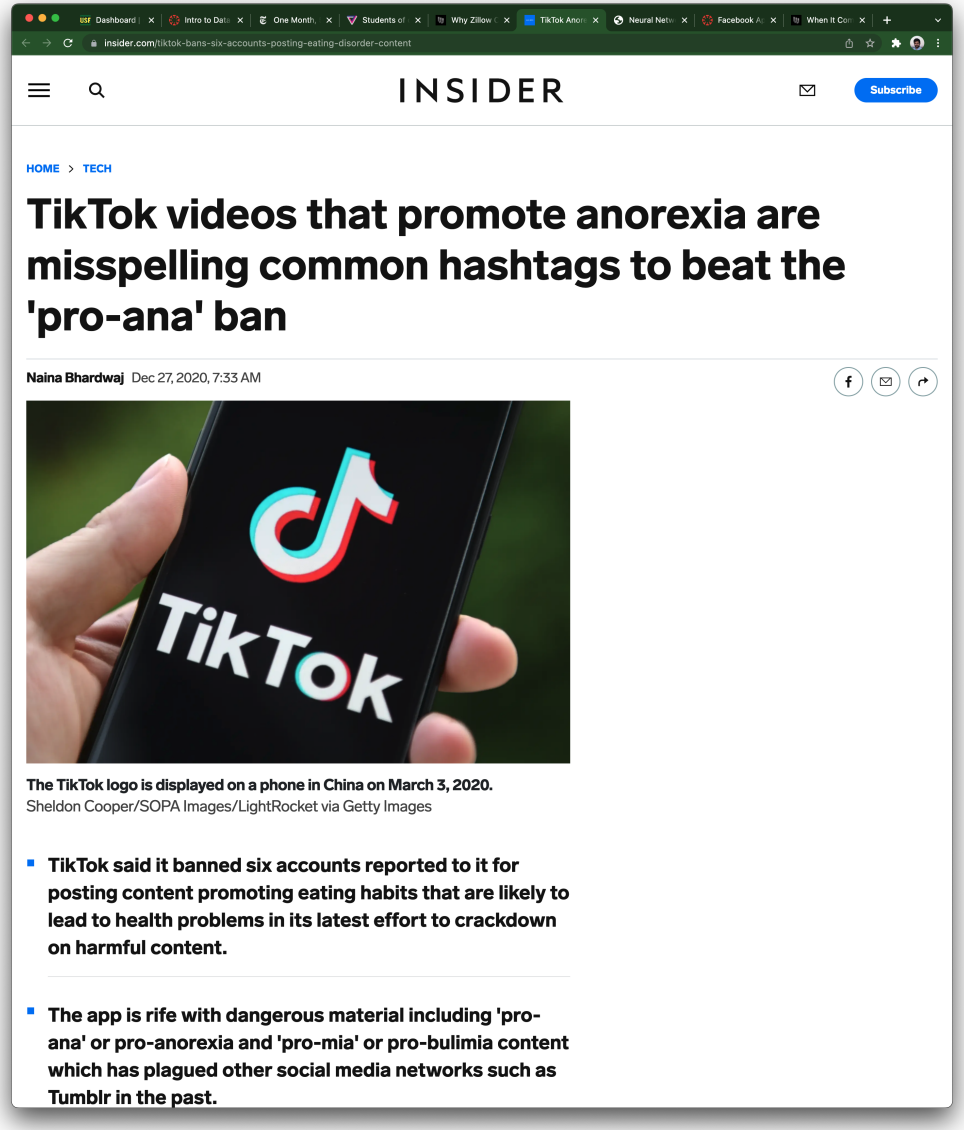
Activists are targeting Russians with open-source "protestware"

Patrick Howell O'Neill

10 Breakthrough Technologies 2022

AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation



insider.com/tiktok-bans-six-accounts-posting-eating-disorder-content

INSIDER

Subscribe

HOME > TECH

TikTok videos that promote anorexia are misspelling common hashtags to beat the 'pro-ana' ban

Naina Bhardwaj Dec 27, 2020, 7:33 AM

The TikTok logo is displayed on a phone in China on March 3, 2020.
Sheldon Cooper/SOPA Images/LightRocket via Getty Images

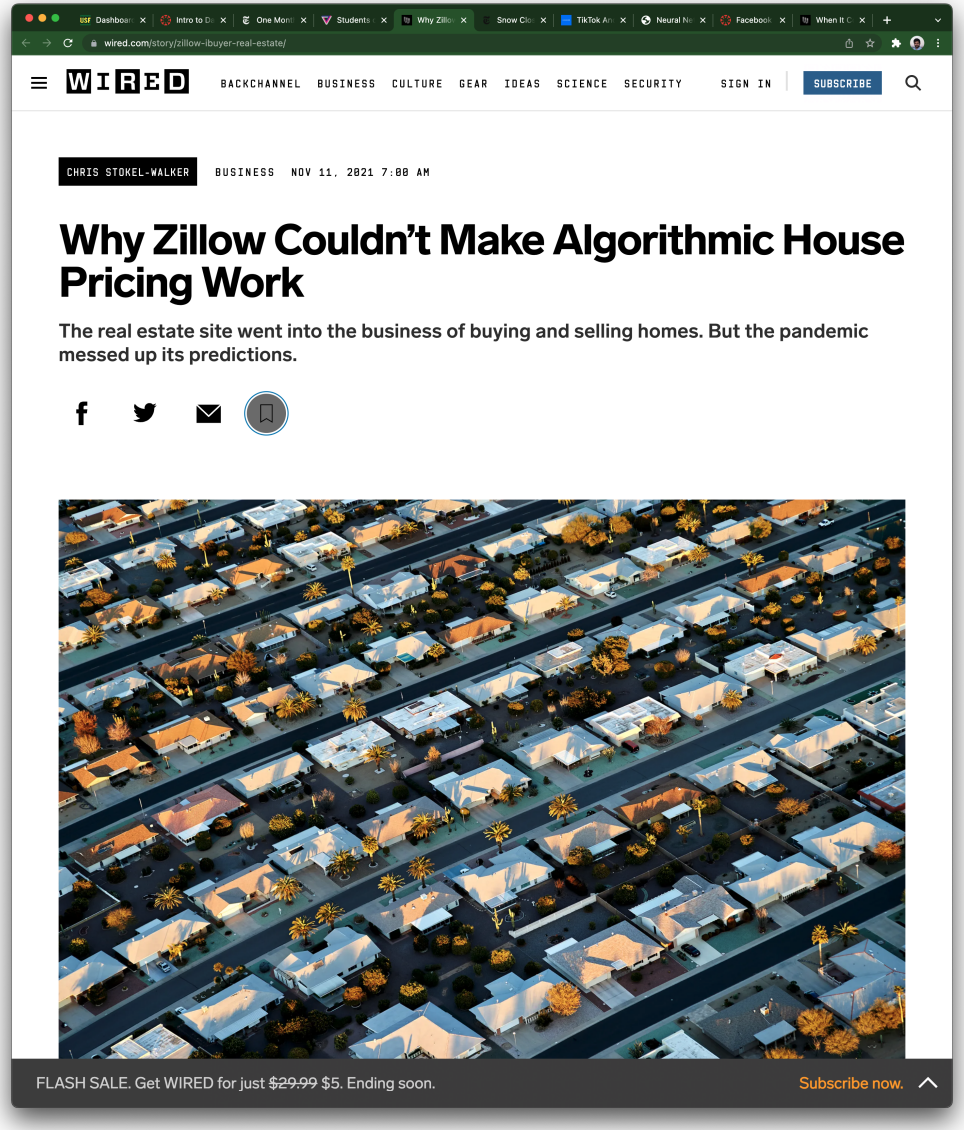
- **TikTok said it banned six accounts reported to it for posting content promoting eating habits that are likely to lead to health problems in its latest effort to crackdown on harmful content.**
- **The app is rife with dangerous material including 'pro-ana' or pro-anorexia and 'pro-mia' or pro-bulimia content which has plagued other social media networks such as Tumblr in the past.**

AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation
- ➔ navigation

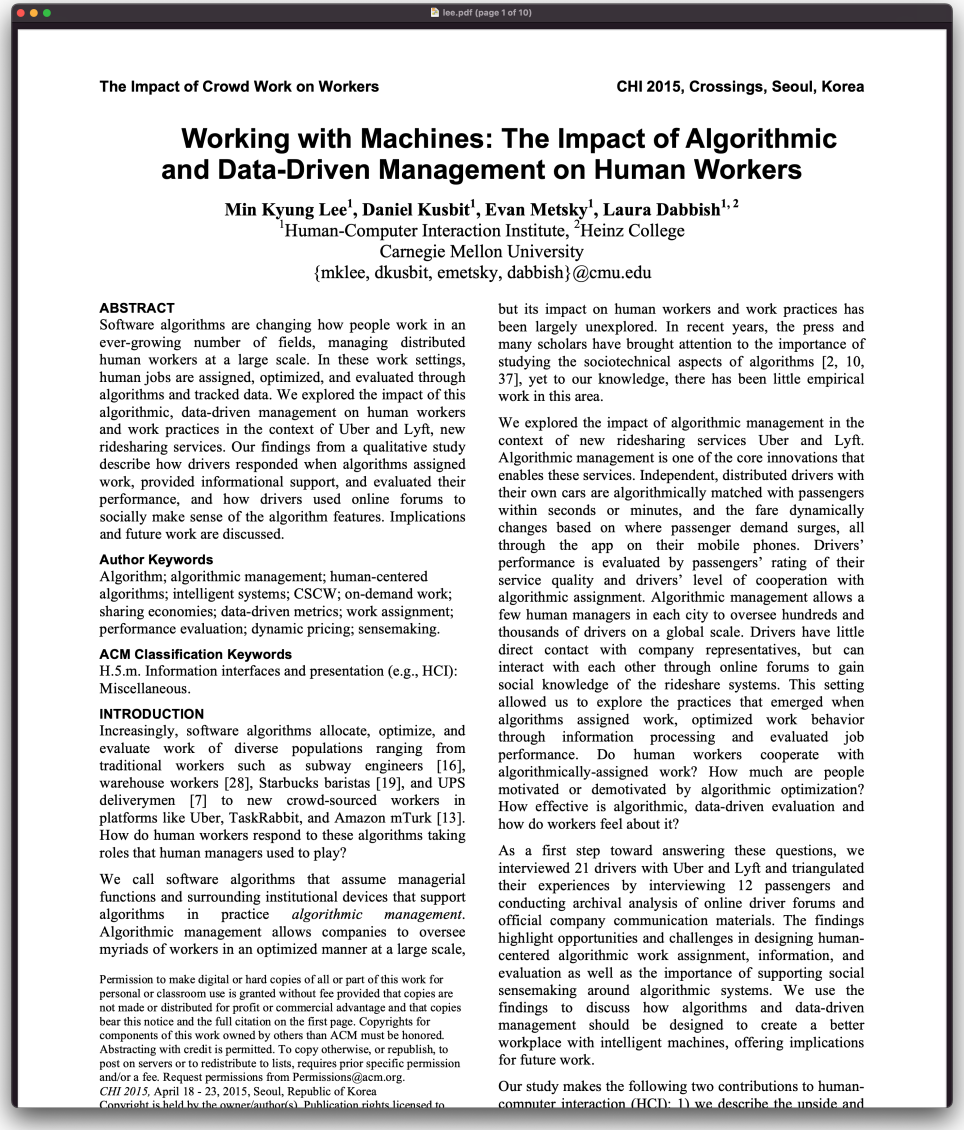
AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation
- ➔ navigation



AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation
- ➔ navigation



lee.pdf (page 1 of 10)

Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers

Min Kyung Lee[1], Daniel Kusbit[1], Evan Metsky[1], Laura Dabbish[1, 2]
[1]Human-Computer Interaction Institute, [2]Heinz College
Carnegie Mellon University
{mklee, dkusbit, emetsky, dabbish}@cmu.edu

ABSTRACT
Software algorithms are changing how people work in an ever-growing number of fields, managing distributed human workers at a large scale. In these work settings, human jobs are assigned, optimized, and evaluated through algorithms and tracked data. We explored the impact of this algorithmic, data-driven management on human workers and work practices in the context of Uber and Lyft, new ridesharing services. Our findings from a qualitative study describe how drivers responded when algorithms assigned work, provided informational support, and evaluated their performance, and how drivers used online forums to socially make sense of the algorithm features. Implications and future work are discussed.

Author Keywords
Algorithm; algorithmic management; human-centered algorithms; intelligent systems; CSCW; on-demand work; sharing economies; data-driven metrics; work assignment; performance evaluation; dynamic pricing; sensemaking.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Increasingly, software algorithms allocate, optimize, and evaluate work of diverse populations ranging from traditional workers such as subway engineers [16], warehouse workers [28], Starbucks baristas [19], and UPS deliverymen [7] to new crowd-sourced workers in platforms like Uber, TaskRabbit, and Amazon mTurk [13]. How do human workers respond to these algorithms taking roles that human managers used to play?

We call software algorithms that assume managerial functions and surrounding institutional devices that support algorithms in practice *algorithmic management*. Algorithmic management allows companies to oversee myriads of workers in an optimized manner at a large scale, but its impact on human workers and work practices has been largely unexplored. In recent years, the press and many scholars have brought attention to the importance of studying the sociotechnical aspects of algorithms [2, 10, 37], yet to our knowledge, there has been little empirical work in this area.

We explored the impact of algorithmic management in the context of new ridesharing services Uber and Lyft. Algorithmic management is one of the core innovations that enables these services. Independent, distributed drivers with their own cars are algorithmically matched with passengers within seconds or minutes, and the fare dynamically changes based on where passenger demand surges, all through the app on their mobile phones. Drivers' performance is evaluated by passengers' rating of their service quality and drivers' level of cooperation with algorithmic assignment. Algorithmic management allows a few human managers in each city to oversee hundreds and thousands of drivers on a global scale. Drivers have little direct contact with company representatives, but can interact with each other through online forums to gain social knowledge of the rideshare systems. This setting allowed us to explore the practices that emerged when algorithms assigned work, optimized work behavior through information processing and evaluated job performance. Do human workers cooperate with algorithmically-assigned work? How much are people motivated or demotivated by algorithmic optimization? How effective is algorithmic, data-driven evaluation and how do workers feel about it?

As a first step toward answering these questions, we interviewed 21 drivers with Uber and Lyft and triangulated their experiences by interviewing 12 passengers and conducting archival analysis of online driver forums and official company communication materials. The findings highlight opportunities and challenges in designing human-centered algorithmic work assignment, information, and evaluation as well as the importance of supporting social sensemaking around algorithmic systems. We use the findings to discuss how algorithms and data-driven management should be designed to create a better workplace with intelligent machines, offering implications for future work.

Our study makes the following two contributions to human-computer interaction (HCI): 1) we describe the upside and

CHI 2015, April 18 - 23, 2015, Seoul, Republic of Korea

AI WILL UNDERSTAND MORE

- → labeling
- → moderation
- → navigation

WHAT DOES IT MEAN TO UNDERSTAND SOMETHING?

UNDERSTANDING

Figuring out what facets you're missing *a priori* is impossible

UNDERSTANDING

Figuring out what facets you're missing *a priori* is impossible

How would you **elicit** this dimension as a designer?

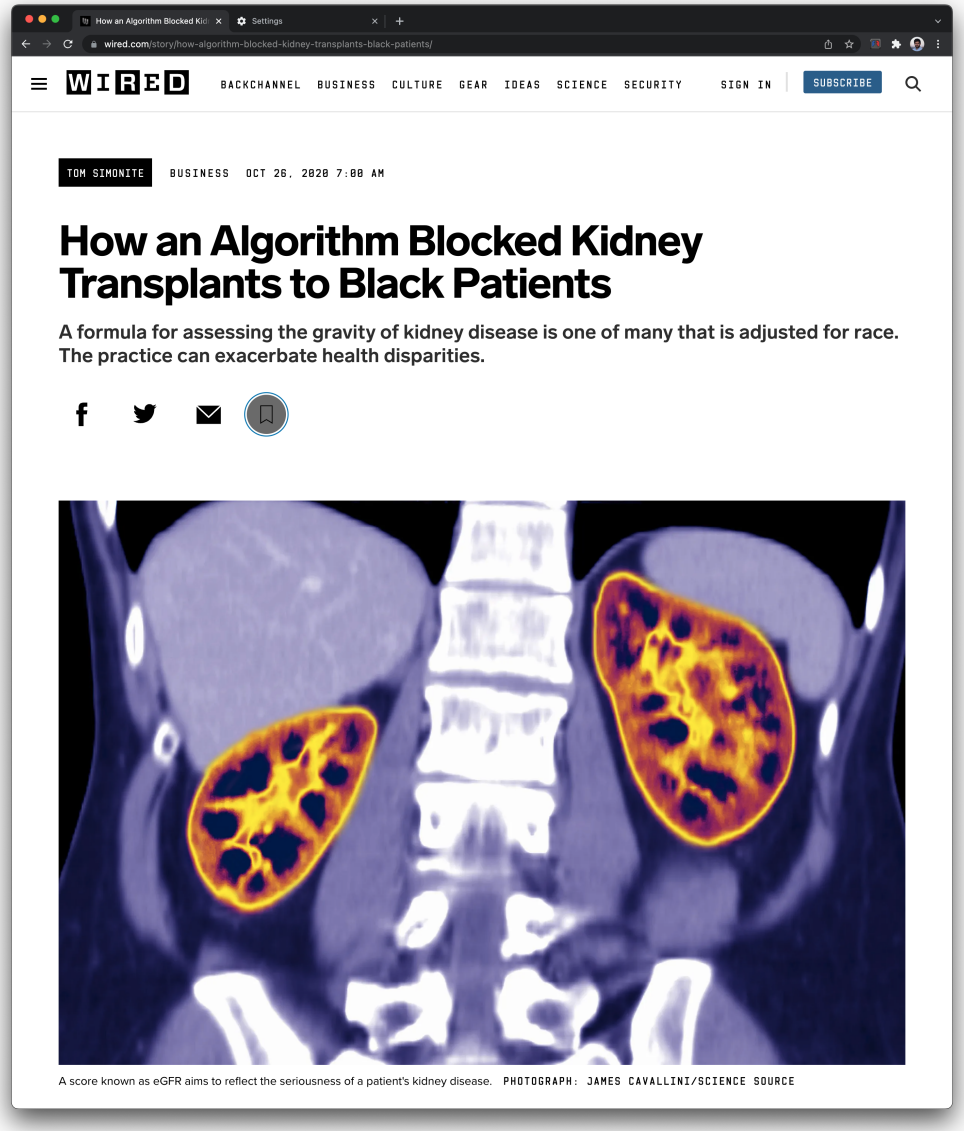
What factors prevent or discourage this kind of approach in corporate settings?

Fairness

AI will be “fairer” or “more objective” than we can be

AI WILL BE FAIRER THAN WE CAN BE

➔ medicine



How an Algorithm Blocked Kid | Settings

wired.com/story/how-algorithm-blocked-kidney-transplants-black-patients/

WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY SIGN IN SUBSCRIBE

TOM SIMONITE BUSINESS OCT 26, 2020 7:00 AM

How an Algorithm Blocked Kidney Transplants to Black Patients

A formula for assessing the gravity of kidney disease is one of many that is adjusted for race. The practice can exacerbate health disparities.

A score known as eGFR aims to reflect the seriousness of a patient's kidney disease. PHOTOGRAPH: JAMES CAVALLINI/SCIENCE SOURCE

AI WILL BE FAIRER THAN WE CAN BE

→ medicine

Publication-78 (2).pdf (page 1 of 15)

ARTICLES
https://doi.org/10.1038/s41591-020-01192-7
nature medicine
Check for updates

An algorithmic approach to reducing unexplained pain disparities in underserved populations

Emma Pierson[1,2], David M. Cutler[3], Jure Leskovec[4], Sendhil Mullainathan[5] ✉ and Ziad Obermeyer[6]

Underserved populations experience higher levels of pain. These disparities persist even after controlling for the objective severity of diseases like osteoarthritis, as graded by human physicians using medical images, raising the possibility that underserved patients' pain stems from factors external to the knee, such as stress. Here we use a deep learning approach to measure the severity of osteoarthritis, by using knee X-rays to predict patients' experienced pain. We show that this approach dramatically reduces unexplained racial disparities in pain. Relative to standard measures of severity graded by radiologists, which accounted for only 9% (95% confidence interval (CI), 3–16%) of racial disparities in pain, algorithmic predictions accounted for 43% of disparities, or 4.7× more (95% CI, 3.2–11.8×), with similar results for lower-income and less-educated patients. This suggests that much of underserved patients' pain stems from factors within the knee not reflected in standard radiographic measures of severity. We show that the algorithm's ability to reduce unexplained disparities is rooted in the racial and socioeconomic diversity of the training set. Because algorithmic severity measures better capture underserved patients' pain, and severity measures influence treatment decisions, algorithmic predictions could potentially redress disparities in access to treatments like arthroplasty.

Pain is widespread and unequally distributed in society. Like many other causes of pain, knee osteoarthritis, which affects 10% of men and 13% of women over 60 years of age in the United States[1], disproportionately affects underserved populations; people of color score far higher on knee pain scales than do white individuals[2–6]. Understanding these racial disparities in pain is important for clinical decision making and public policy but also for understanding pain disparities for a variety of other medical problems[7,8].

Two explanations for these disparities have been proposed. First, underserved patients might have more severe osteoarthritis within the knee. Alternatively, underserved patients could have more aggravating factors external to the knee. For example, the same physical ailments in different populations can produce very different experienced pain due to life stress, social isolation or other factors[7–9]. These two explanations have very different treatment implications: psychosocial interventions target causes external to the knee, whereas physical therapy, medication and orthopedic procedures address causes within the knee[10–12].

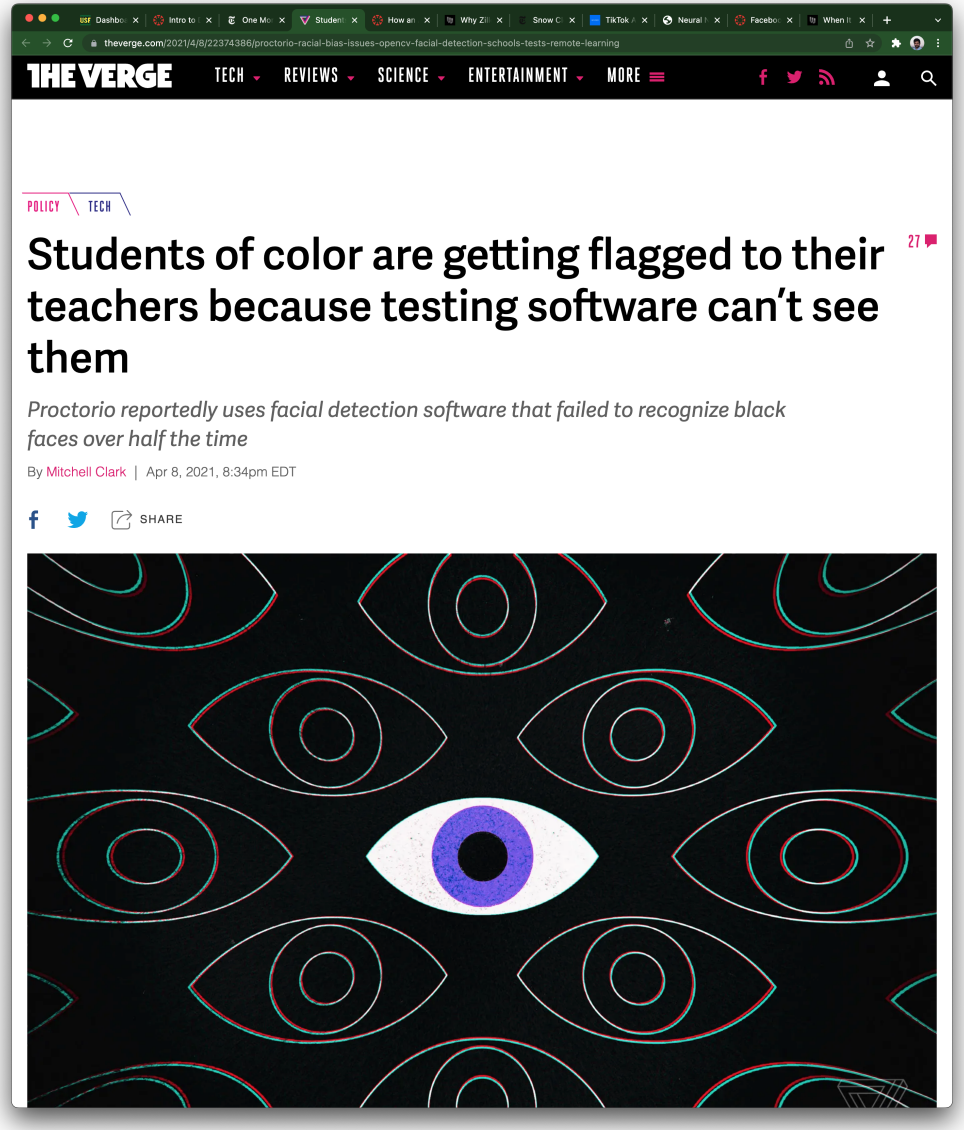
Research to date has indirectly implicated factors external to the knee. Methodologically, this is demonstrated by defining an objective measure of osteoarthritis severity based on knee X-rays and then measuring the extent of pain disparities that remain after adjusting for severity. Typically, large differences in pain remain even after adjustment[2–4,13]. For example, even though Black patients have more severe osteoarthritis based on standard radiographic measures (Kellgren–Lawrence grade (KLG)), adjusting for KLG only slightly decreased measured Black–white disparities in pain[3,13]. These findings that pain disparities remain even when adjusting for radiographic osteoarthritis severity, however, depend heavily on how severity is measured. The relationship between radiographic severity and pain is debated. Many patients with mild or no disease as measured by radiographic severity suffer pain, and many patients with structural damage on X-ray or even magnetic resonance imaging (MRI) experience no or very little pain[14–16]. Standard radiographic measures such as KLG, developed decades ago in white British populations, might miss physical causes of pain in people of color[17,18]; further, there are known racial and socioeconomic biases in how a patient's pain is perceived by observers[19,20]. If the pain experienced by underserved populations is caused by objective factors missing from current measures, a range of painful, treatable knee ailments would be misattributed to factors external to the knee.

In this paper, we use a machine-learning approach to discriminate between the 'within the knee' and 'external to the knee' hypotheses. We produce a new algorithmic measure of osteoarthritis severity from radiographs alone. We use a dataset of knee radiographs from a diverse sample of 4,172 patients in the United States who had or were at high risk of developing knee osteoarthritis. As part of an NIH-funded study[21], bilateral fixed flexion knee radiographs were obtained and scored by radiologists on summary measures of radiographic severity (for example, KLG) and other objective features (for example, osteophytes and joint space narrowing (JSN)). Patients also reported a knee-specific pain score (Knee injury and Osteoarthritis Outcome Score (KOOS)), derived from a multi-item survey on pain experienced during various activities (for example, fully straightening the knee[22]).

Summary statistics of the 4,172 participants, who generated 36,369 observations (one for each knee at each time point) are provided in Table 1. Black patients had substantially higher pain levels across knees and time points compared with non-Black patients (97% of whom were white). Black patients experienced severe pain 58% of the time (KOOS ≤ 86.1, a standard threshold for severe pain[23]), compared with 38% for patients overall (*P* value for racial difference, <0.001). The median Black patient had worse pain than 75% of non-Black patients. Black patients had a pain score that was 10.6 KOOS points higher than that of non-Black patients

AI WILL BE FAIRER THAN WE CAN BE

→ medicine



theverge.com/2021/4/8/22374386/proctorio-racial-bias-issues-opencv-facial-detection-schools-tests-remote-learning

THE VERGE TECH REVIEWS SCIENCE ENTERTAINMENT MORE

POLICY TECH

27

Students of color are getting flagged to their teachers because testing software can't see them

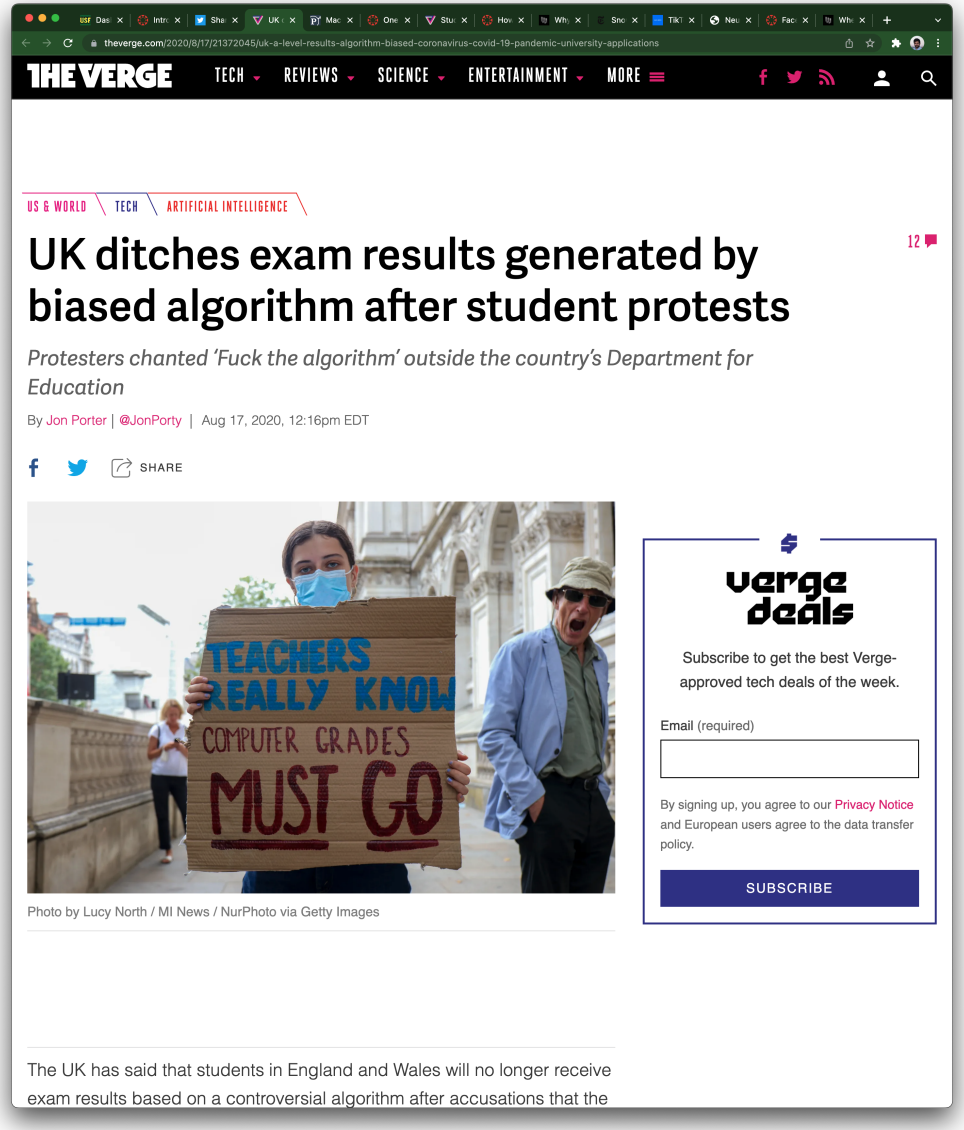
Proctorio reportedly uses facial detection software that failed to recognize black faces over half the time

By Mitchell Clark | Apr 8, 2021, 8:34pm EDT

SHARE

AI WILL BE FAIRER THAN WE CAN BE

- ➔ medicine
- ➔ education



theverge.com/2020/8/17/21372045/uk-a-level-results-algorithm-biased-coronavirus-covid-19-pandemic-university-applications

THE VERGE TECH REVIEWS SCIENCE ENTERTAINMENT MORE

US & WORLD | TECH | ARTIFICIAL INTELLIGENCE

12

UK ditches exam results generated by biased algorithm after student protests

Protesters chanted 'Fuck the algorithm' outside the country's Department for Education

By Jon Porter | @JonPorty | Aug 17, 2020, 12:16pm EDT

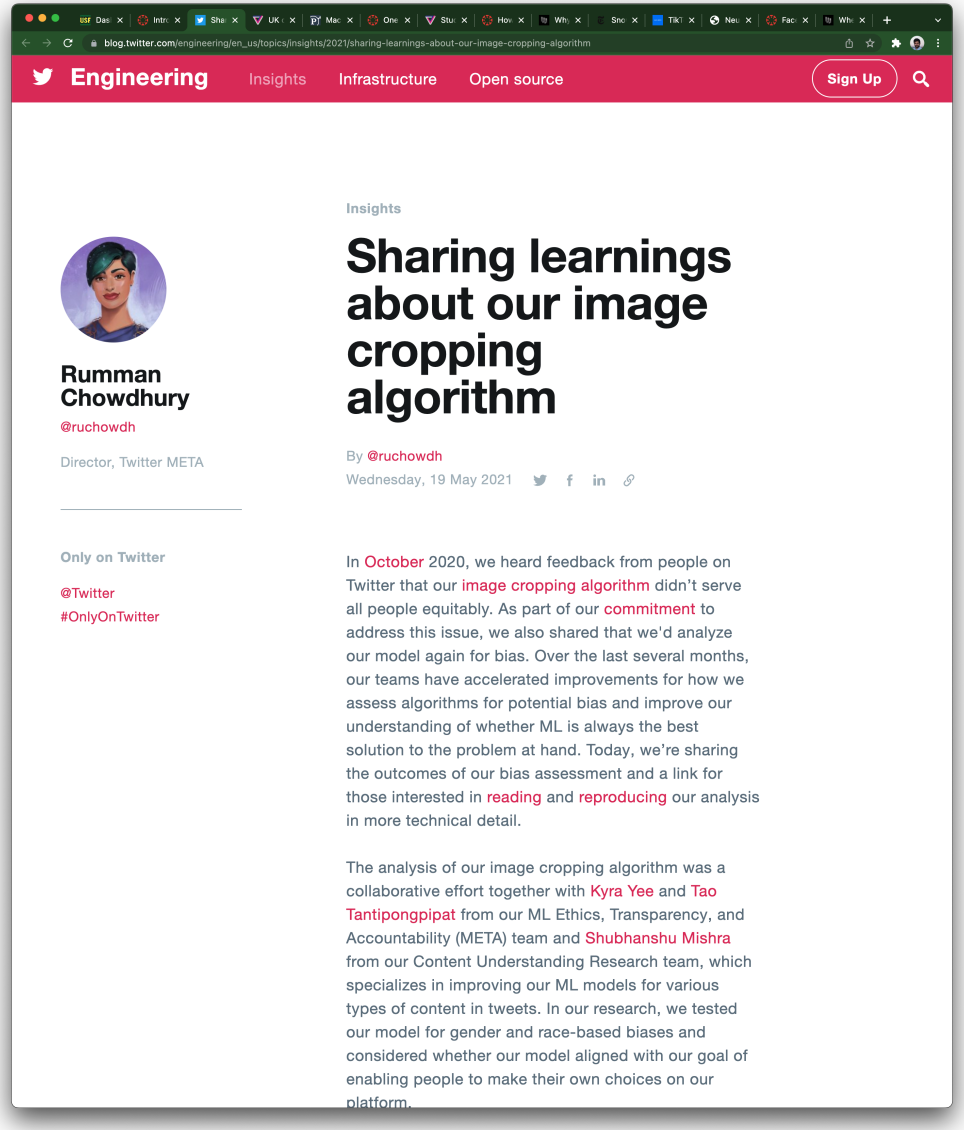
SHARE

Photo by Lucy North / MI News / NurPhoto via Getty Images

The UK has said that students in England and Wales will no longer receive exam results based on a controversial algorithm after accusations that the

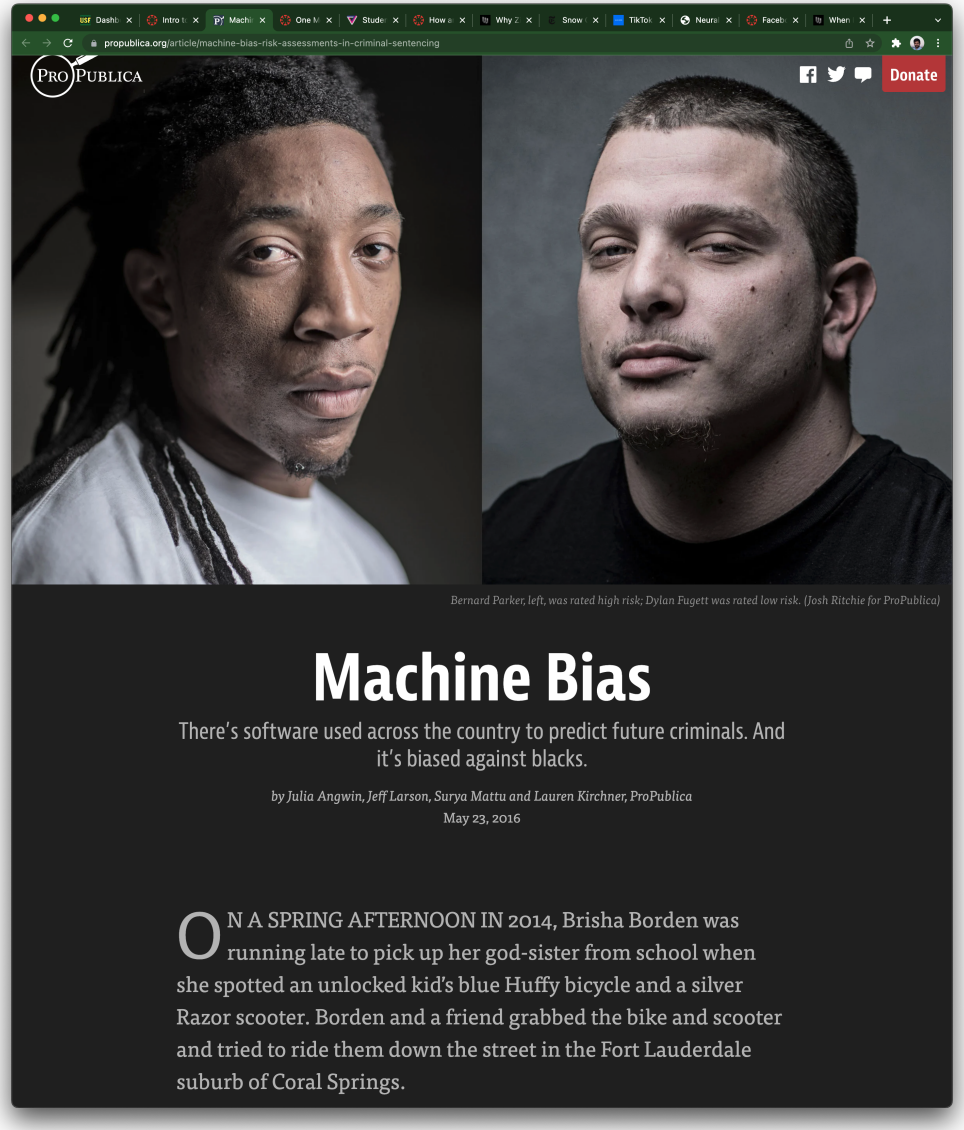
AI WILL BE FAIRER THAN WE CAN BE

- ➔ medicine
- ➔ education
- ➔ social



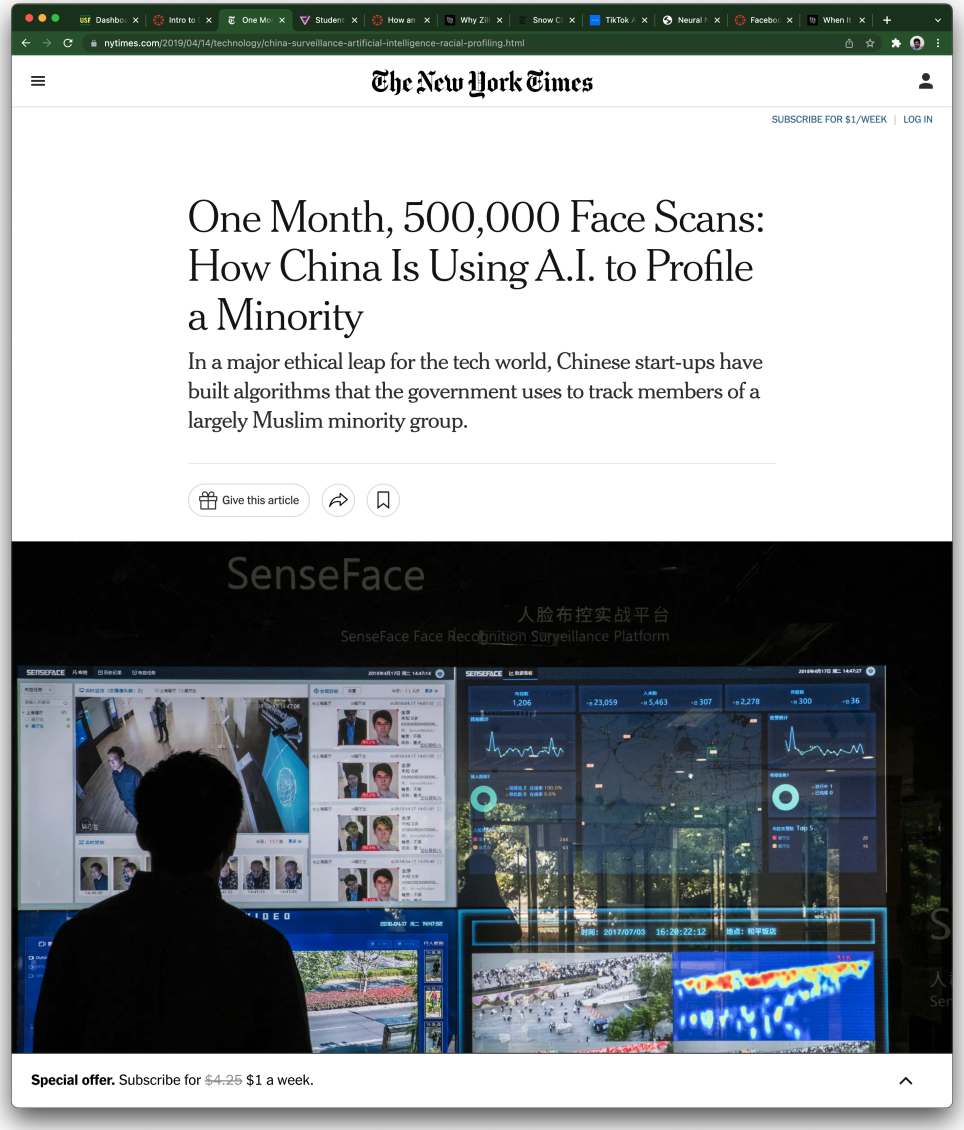
AI WILL BE FAIRER THAN WE CAN BE

- ➔ medicine
- ➔ education
- ➔ social



AI WILL BE FAIRER THAN WE CAN BE

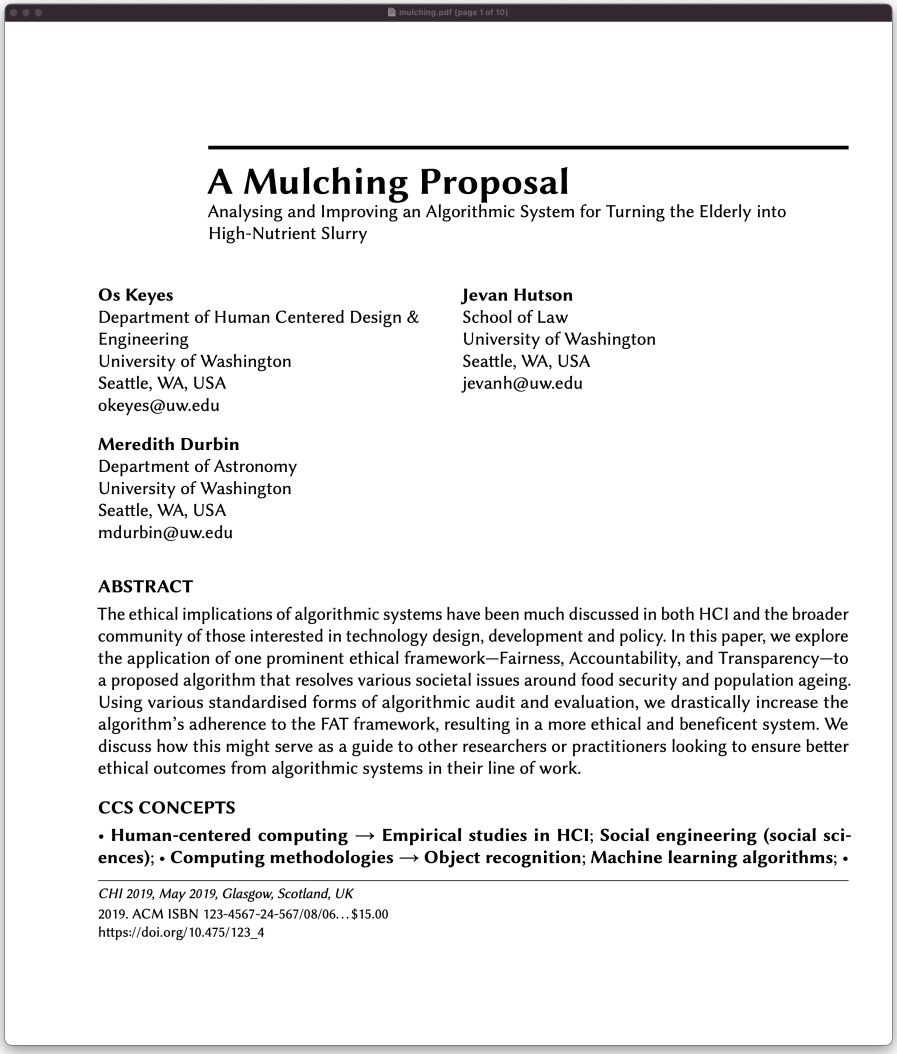
- ➔ medicine
- ➔ education
- ➔ social



WHAT IS FAIRNESS?

FAIRNESS

A Mulching Proposal



mulching.pdf (page 1 of 10)

A Mulching Proposal

Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry

Os Keyes
Department of Human Centered Design & Engineering
University of Washington
Seattle, WA, USA
okeyes@uw.edu

Jevan Hutson
School of Law
University of Washington
Seattle, WA, USA
jevanh@uw.edu

Meredith Durbin
Department of Astronomy
University of Washington
Seattle, WA, USA
mdurbin@uw.edu

ABSTRACT

The ethical implications of algorithmic systems have been much discussed in both HCI and the broader community of those interested in technology design, development and policy. In this paper, we explore the application of one prominent ethical framework—Fairness, Accountability, and Transparency—to a proposed algorithm that resolves various societal issues around food security and population ageing. Using various standardised forms of algorithmic audit and evaluation, we drastically increase the algorithm's adherence to the FAT framework, resulting in a more ethical and beneficent system. We discuss how this might serve as a guide to other researchers or practitioners looking to ensure better ethical outcomes from algorithmic systems in their line of work.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Social engineering (social sciences)**; • **Computing methodologies** → **Object recognition**; **Machine learning algorithms**; •

CHI 2019, May 2019, Glasgow, Scotland, UK
2019. ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

FAIRNESS

The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning

FAIRNESS

The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning

- literacy tests
- redlining
- proxies, like
  - zip code
  - last name

FAIRNESS

The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning

"…there are important cases where even protected group membership itself should be explicitly taken into account to make equitable decisions."

FAIRNESS

To varying degrees, we are **mediators** of “fair” or “equitable” or “just” outcomes.

FAIRNESS

To varying degrees, we are **mediators** of "fair" or "equitable" or "just" outcomes.

How can you identify **stakeholders** who might care how the system adjudicates issues?

What is your responsibility regarding a system **after** deployment?