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

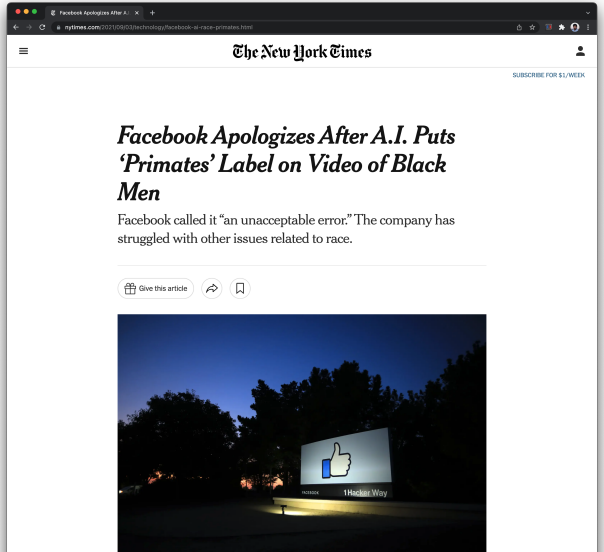
- AI can “understand” the world better than we can
- AI can make “fairer” decisions than we can

Understanding

AI WILL “UNDERSTAND” BETTER THAN WE CAN

AI WILL UNDERSTAND MORE

→ labeling

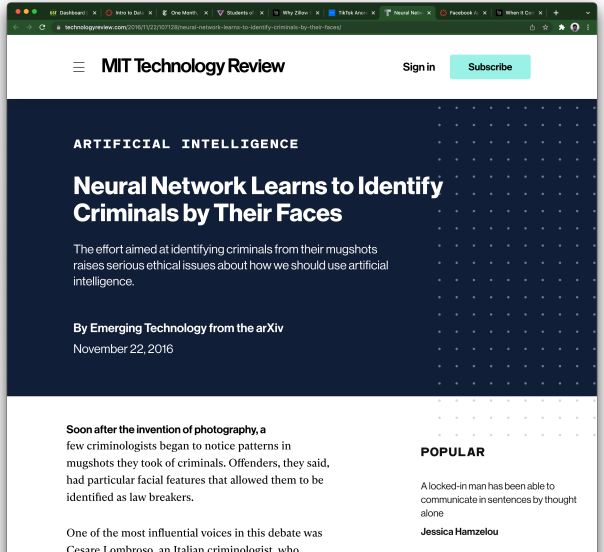


AI WILL UNDERSTAND MORE

→ labeling

AI WILL UNDERSTAND MORE

→ labeling



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

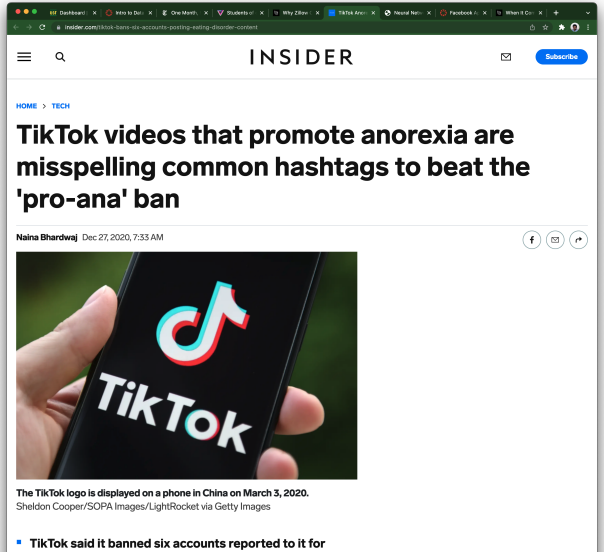
POPULAR

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

Jessica Hamzelou

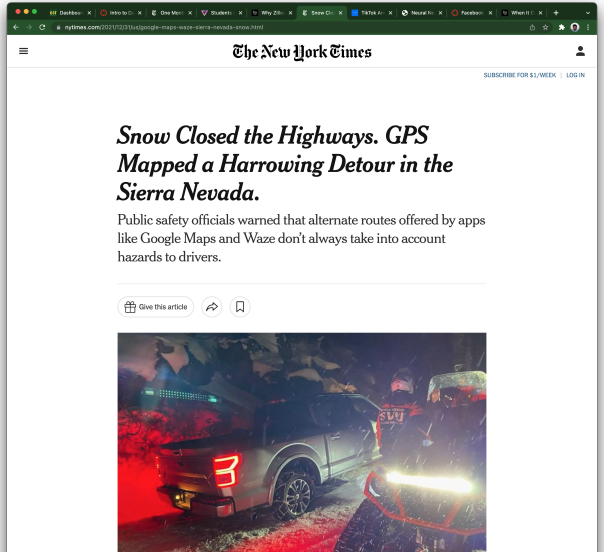
AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation



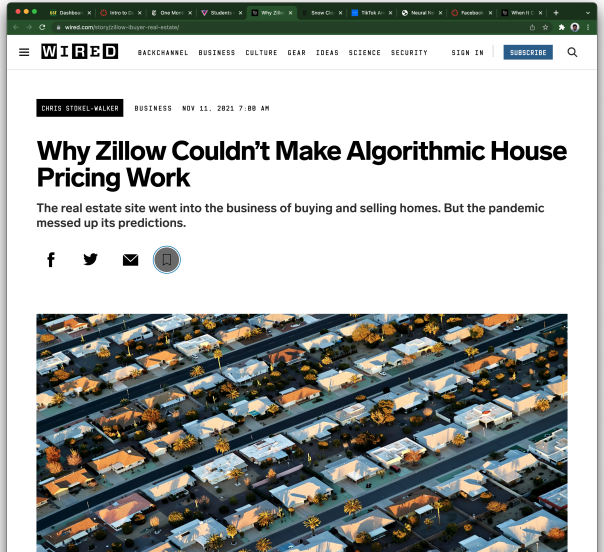
AI WILL UNDERSTAND MORE

- labeling
- moderation
- navigation



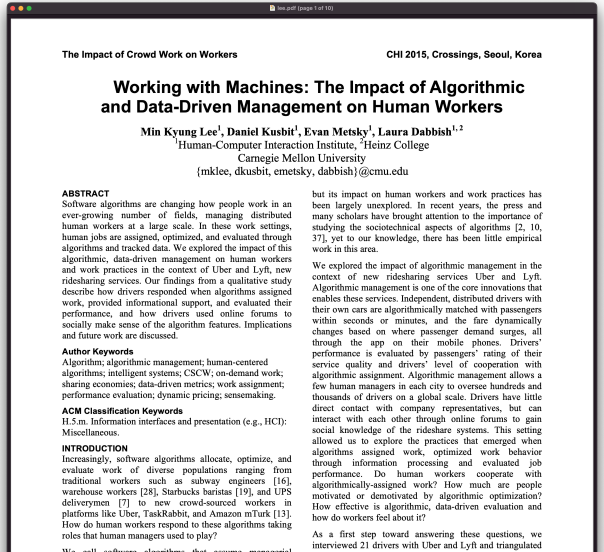
AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation
- ➔ navigation



AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation
- ➔ navigation



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?

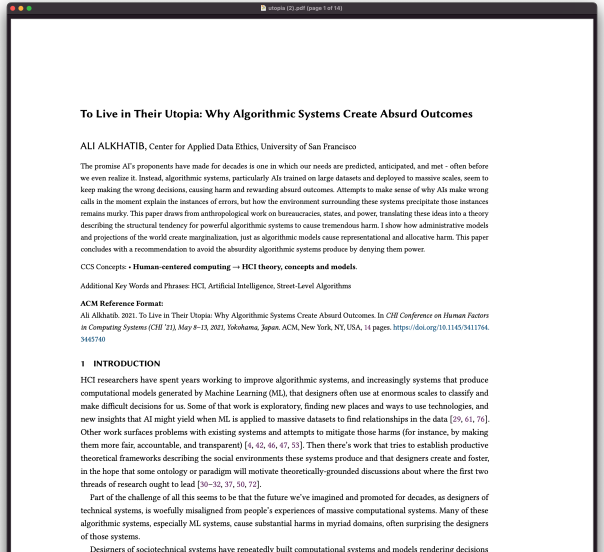
but its impact on human workers and work practices has been largely unexplored. In recent years, the press and many scholars have brought attention to the importance of studying the sociotechnical aspects of algorithms [2, 10, 37], yet to our knowledge, there has been little empirical work in this area.

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

AI WILL UNDERSTAND MORE

- ➔ labeling
- ➔ moderation
- ➔ navigation



utopia (2).pdf (page 1 of 14)

To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes

ALI ALKHATIB, Center for Applied Data Ethics, University of San Francisco

The promise AI's proponents have made for decades is one in which our needs are predicted, anticipated, and met - often before we even realize it. Instead, algorithmic systems, particularly AIs trained on large datasets and deployed to massive scales, seem to keep making the wrong decisions, causing harm and rewarding absurd outcomes. Attempts to make sense of why AIs make wrong calls in the moment explain the instances of errors, but how the environment surrounding these systems precipitate those instances remains murky. This paper draws from anthropological work on bureaucracies, states, and power, translating these ideas into a theory describing the structural tendency for powerful algorithmic systems to cause tremendous harm. I show how administrative models and projections of the world create marginalization, just as algorithmic models cause representational and allocative harm. This paper concludes with a recommendation to avoid the absurdity algorithmic systems produce by denying them power.

CCS Concepts: • **Human-centered computing → HCI theory, concepts and models.**

Additional Key Words and Phrases: HCI, Artificial Intelligence, Street-Level Algorithms

ACM Reference Format:
Ali Alkhatib. 2021. To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445740

1 INTRODUCTION

HCI researchers have spent years working to improve algorithmic systems, and increasingly systems that produce computational models generated by Machine Learning (ML), that designers often use at enormous scales to classify and make difficult decisions for us. Some of that work is exploratory, finding new places and ways to use technologies, and new insights that AI might yield when ML is applied to massive datasets to find relationships in the data [29, 61, 76]. Other work surfaces problems with existing systems and attempts to mitigate those harms (for instance, by making them more fair, accountable, and transparent) [4, 42, 46, 47, 53]. Then there's work that tries to establish productive theoretical frameworks describing the social environments these systems produce and that designers create and foster, in the hope that some ontology or paradigm will motivate theoretically-grounded discussions about where the first two threads of research ought to lead [30–32, 37, 50, 72].

Part of the challenge of all this seems to be that the future we've imagined and promoted for decades, as designers of technical systems, is woefully misaligned from people's experiences of massive computational systems. Many of these algorithmic systems, especially ML systems, cause substantial harms in myriad domains, often surprising the designers of those systems.

Designers of sociotechnical systems have repeatedly built computational systems and models rendering decisions

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?

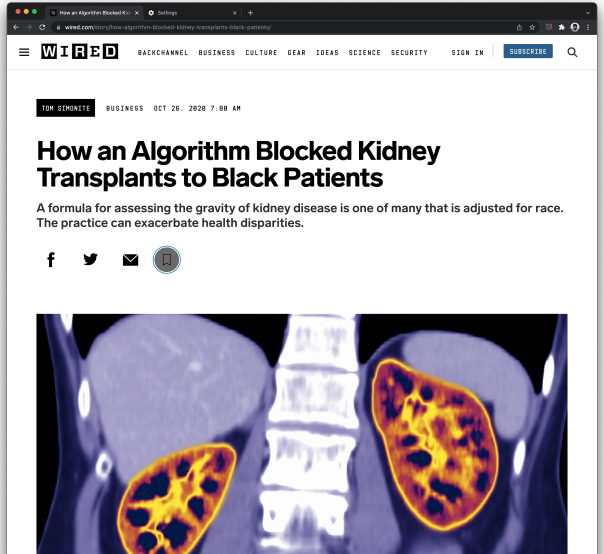
What factors prevent or discourage this kind of approach in corporate settings?

Fairness

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

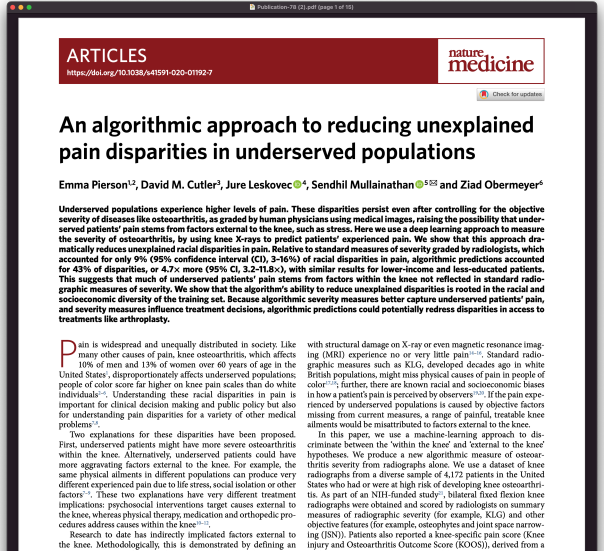
AI WILL BE FAIRER THAN WE CAN BE

→ medicine



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[4,5] ✉ 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[2–4]. 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[5,6].

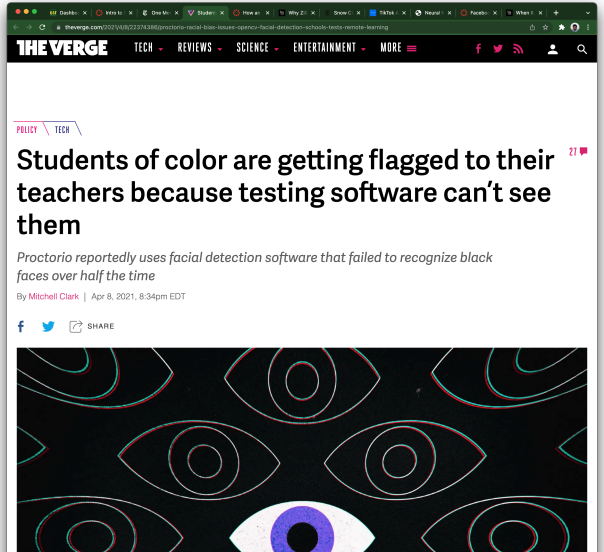
Two explanations for these disparities have been proposed. First, underserved patients might have more severe osteoarthritis within the knee. Alternatively, underserved patients could have more aggravating factors external to the knee. For example, the same physical ailments in different populations 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 with structural damage on X-ray or even magnetic resonance imaging (MRI) experience no or very little pain[13–19]. 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[20,21]. 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[22], 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

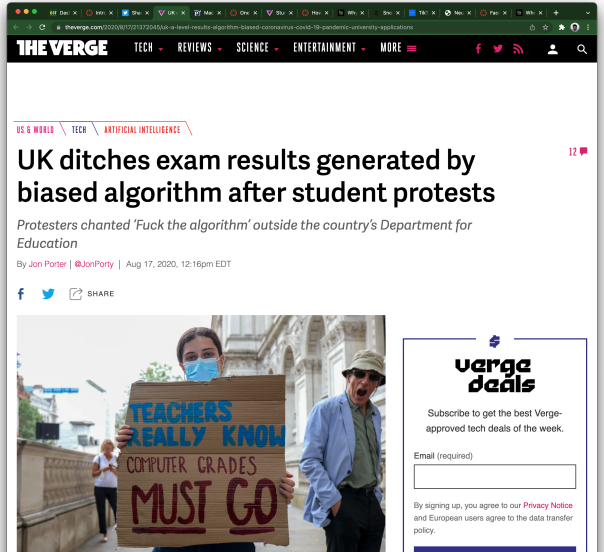
AI WILL BE FAIRER THAN WE CAN BE

→ medicine



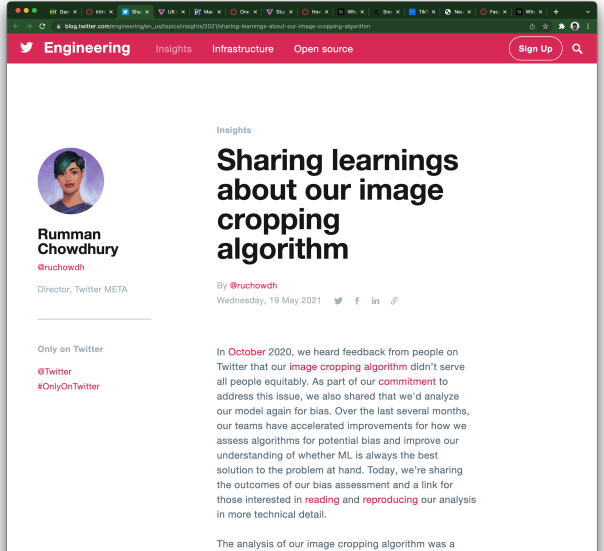
AI WILL BE FAIRER THAN WE CAN BE

- ➔ medicine
- ➔ education



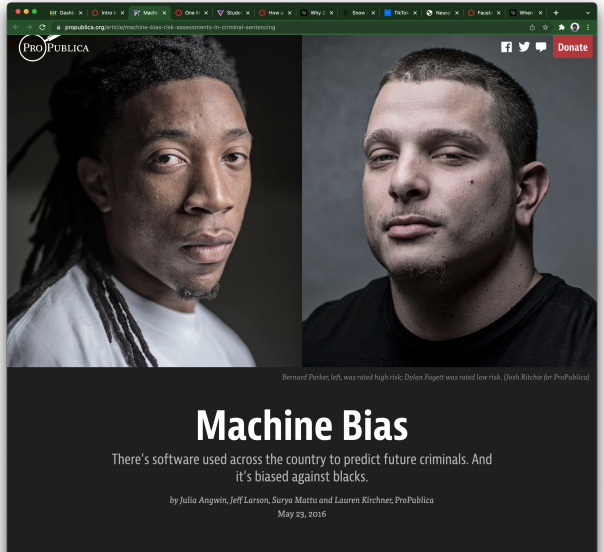
AI WILL BE FAIRER THAN WE CAN BE

- ➔ medicine
- ➔ education
- ➔ social



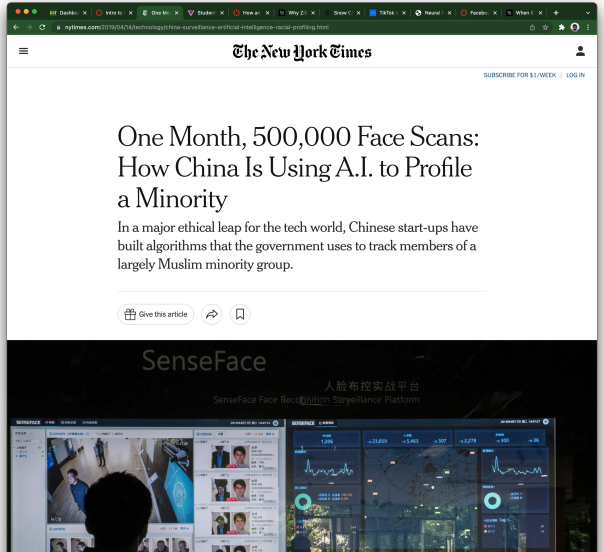
AI WILL BE FAIRER THAN WE CAN BE

- ➔ medicine
- ➔ education
- ➔ social



AI WILL BE FAIRER THAN WE CAN BE

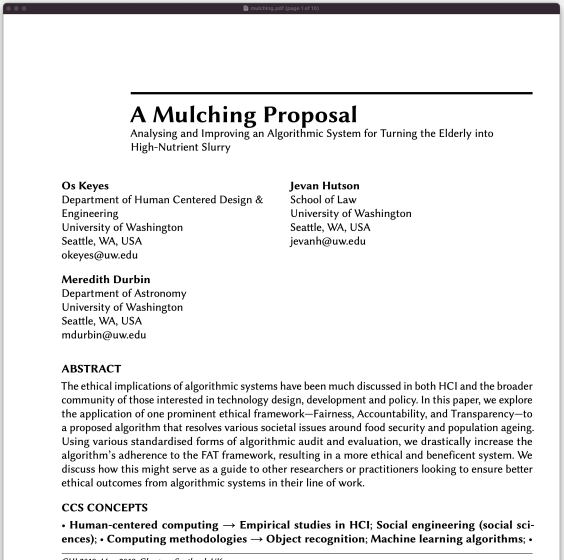
- ➔ medicine
- ➔ education
- ➔ social



WHAT IS FAIRNESS?

FAIRNESS

A Mulching Proposal



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**; •

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?