Of Humans and AI: Social and Ethical Implications of AI for Decision-Making

CORIOLIS SEMINAR, NOVEMBER 6 2025

SOIZIC PÉNICAUD – SOIZIC.PENICAUD@SCIENCESPO.FR

Background in Law and Social Sciences

Former member of Etalab (French public service)

Now:

- Independent consultant and researcher on AI in the public sector (with qualitative methods)
- Cofounder, Odap.fr
- Lecturer, Sciences Po Paris

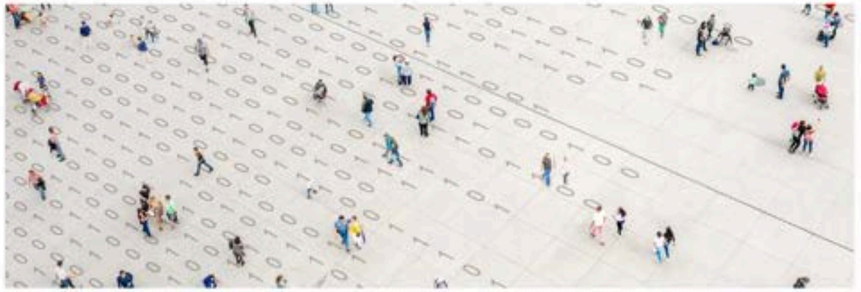
What human choices and assumptions are embedded in AI systems?*

What human choices and assumptions are embedded in AI systems?*

*AI understood as a broad range of technologies that generate outputs such as predictions, content, recommendations, or decisions.

AI isn’t objective: automated decision-making in social protection

LIGHTHOUSE REPORTS



SURVEILLANCE | SUSPICION MACHINES

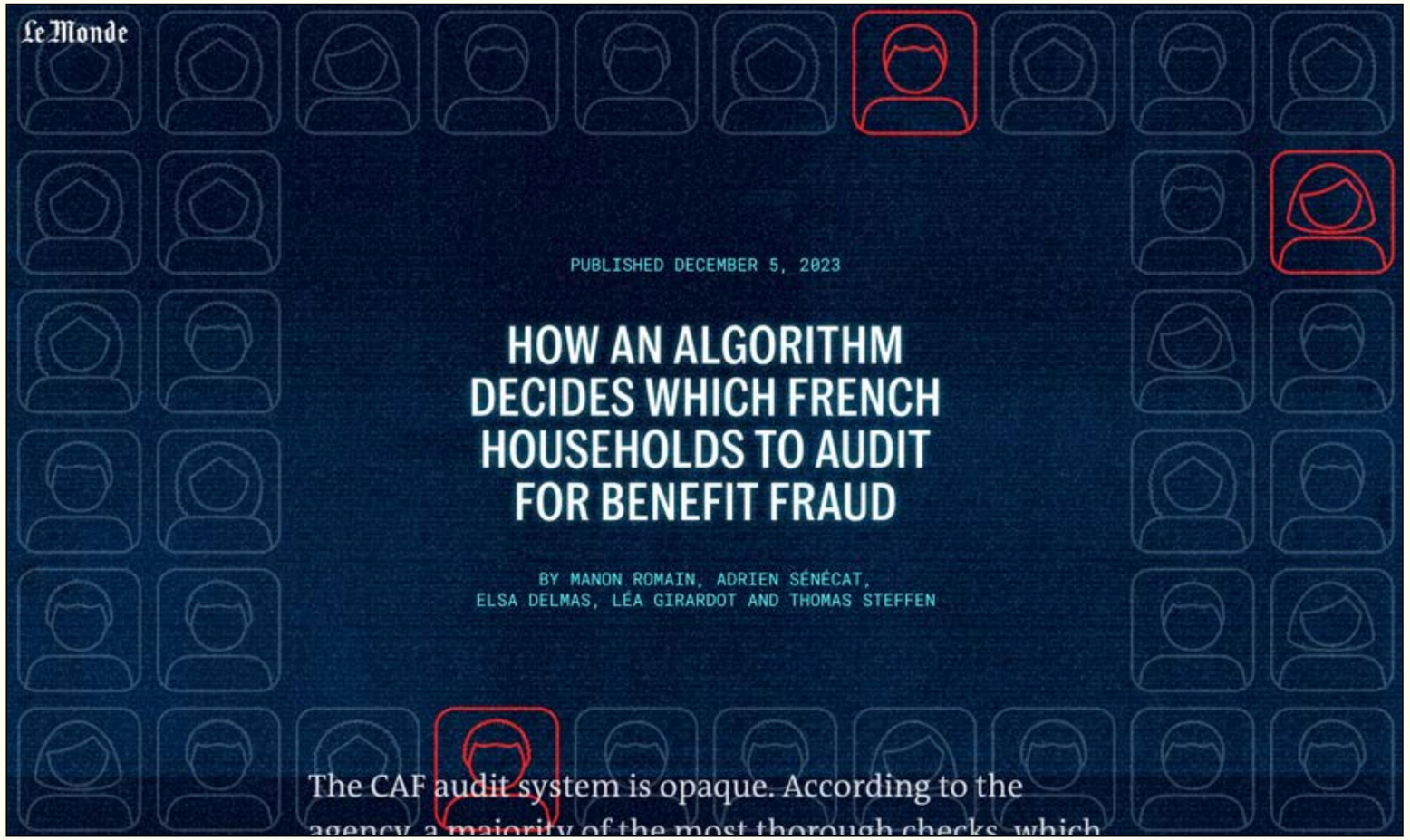
France's Digital Inquisition

Taking apart the secretive fraud detection algorithm that scores half of France's population but pursues the most vulnerable.

In 2022, Juliette, a single mother on welfare, received money from her family to visit her critically ill father. A few months after her father died, a fraud investigator from France's social security agency, CNAF, knocked on her door. The investigation determined that she owed thousands of euros that would be deducted from her monthly welfare payments.

What Juliette did not know at the time was that she was one of hundreds of thousands of people on welfare in France being flagged by an algorithm.

For more than a decade and without any public consultation, CNAF has deployed machine learning at a massive scale in a hunt for



https://git.laquadrature.net/la-quadrature-du-net/algo-et-controle/caf

Some of the criteria leading to a higher score

Working while receiving a disability pension

Having a child who is over 19 years old

Undergoing a change in family situation

Having an irregular income

Some of the criteria leading to a higher score

Working while receiving a disability pension

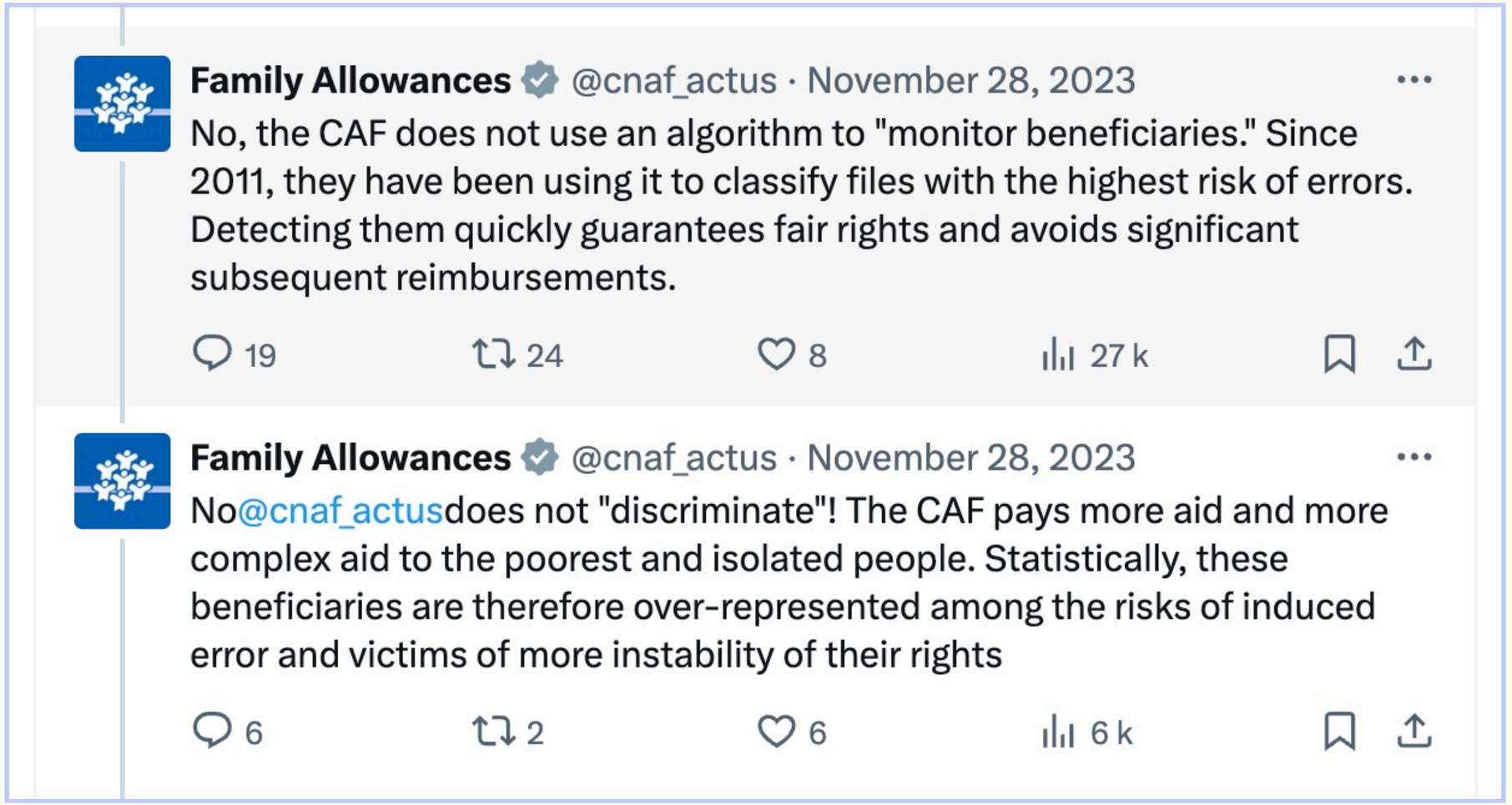
Having a child who is over 19 years old

Undergoing a change in family situation

Having an irregular income

Single parents account for 36% of at-home inspections while making up only 16% of households receiving benefits. (*Le Monde*, 2023)

What the French social protection agency says



Family Allowances @cnaf_actus · November 28, 2023

No, the CAF does not use an algorithm to "monitor beneficiaries." Since 2011, they have been using it to classify files with the highest risk of errors. Detecting them quickly guarantees fair rights and avoids significant subsequent reimbursements.

19 24 8 27 k

Family Allowances @cnaf_actus · November 28, 2023

No@cnaf_actusdoes not "discriminate"! The CAF pays more aid and more complex aid to the poorest and isolated people. Statistically, these beneficiaries are therefore over-represented among the risks of induced error and victims of more instability of their rights

6 2 6 6 k

“Detecting errors quickly guarantees fair rights” -yet…

…the **choice** of target: overpayments

“Detecting errors quickly guarantees fair rights” -yet…

“Detecting errors quickly guarantees fair rights” -yet…

…the **choice** of target: overpayments

…the **choice** to focus on post-hoc detection rather than prevention

“Detecting errors quickly guarantees fair rights” -yet…

…the **choice** of target: overpayments

…the **choice** to focus on post-hoc detection rather than prevention

This latter problem can’t necessarily be solved by technology!

“Statistically, these beneficiaries are overrepresented among the risks of induced error”

“Statistically, these beneficiaries are overrepresented among the risks of induced error”

… but statistically accurate doesn’t mean legal.

Using AI/automated decision-making **obfuscates human choices**, under a veneer of objectivity and statistical accuracy.

Fairness and functionality: the case of COMPAS

PRO PUBLICA

Donate

Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

« The impossibility of fairness »

Sections

The Washington Post
Democracy Dies in Darkness

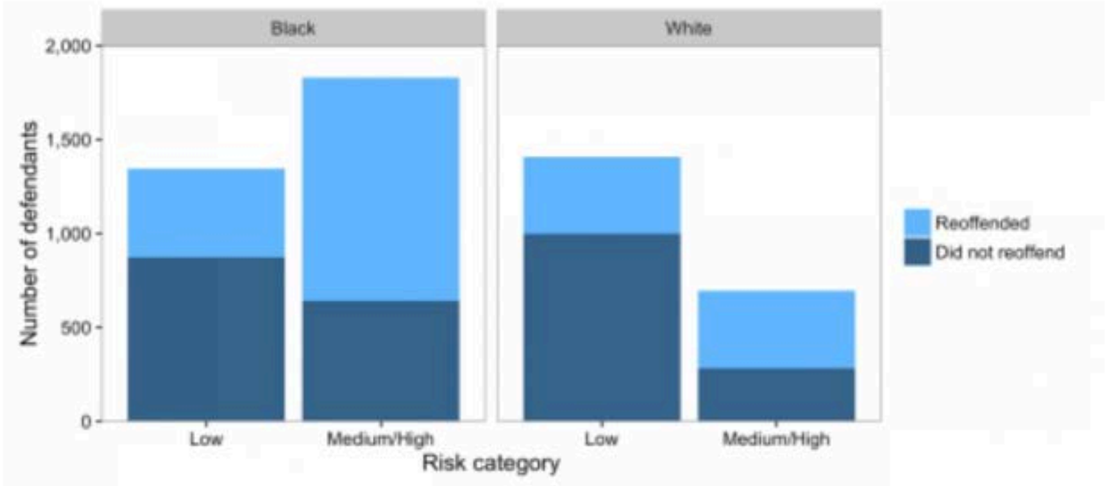
Get one year for $40

Sign in

This article is more than **1 year old**

Monkey Cage

A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.



Distribution of defendants across risk categories by race. Black defendants reoffended at a higher rate than whites, and accordingly, a higher proportion of black defendants are deemed medium or high risk. As a result, blacks who do not reoffend are also more likely to be classified higher risk than whites who do not reoffend.

Fair... for whom? Based on what? Developers **choose** their fairness metric and the tradeoffs they want to make.

ScienceAdvances

Current Issue First release papers Archive Ab

RESEARCH ARTICLE | RESEARCH METHODS

The accuracy, fairness, and limits of predicting recidivism

JULIA DRESSEL AND HANY FARID Authors Info & Affiliations

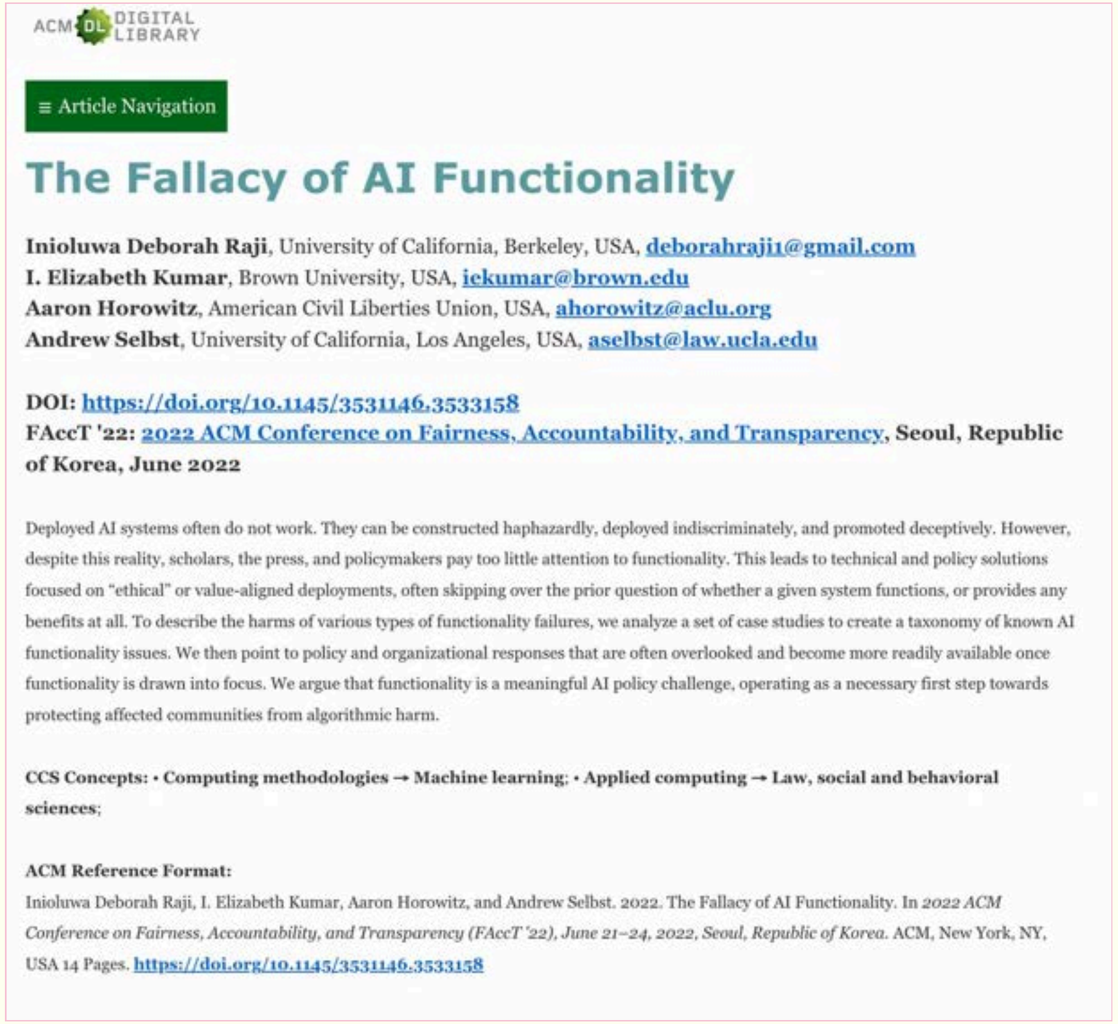
SCIENCE ADVANCES • 17 Jan 2018 • Vol 4, Issue 1 • DOI: 10.1126/sciadv.aao5580

61 977 559

Abstract

Algorithms for predicting recidivism are commonly used to assess a criminal defendant's likelihood of committing a crime. These predictions are used in pretrial, parole, and sentencing decisions. Proponents of these systems argue that big data and advanced machine learning make these analyses more accurate and less biased than humans. We show, however, that the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise. In addition, despite COMPAS's collection of 137 features, the same accuracy can be achieved with a simple linear classifier with only two features.

“The fallacy of AI functionality”



ACM DIGITAL LIBRARY

≡ Article Navigation

The Fallacy of AI Functionality

Inioluwa Deborah Raji, University of California, Berkeley, USA, deborahraji1@gmail.com
I. Elizabeth Kumar, Brown University, USA, iekumar@brown.edu
Aaron Horowitz, American Civil Liberties Union, USA, ahorowitz@aclu.org
Andrew Selbst, University of California, Los Angeles, USA, aselbst@law.ucla.edu

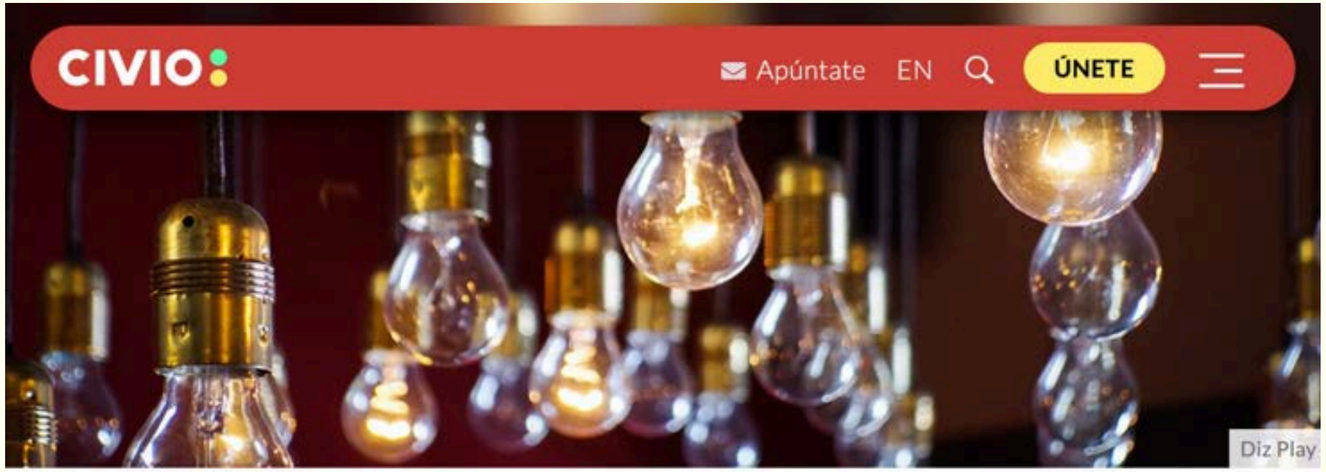
DOI: https://doi.org/10.1145/3531146.3533158
FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, **Seoul, Republic of Korea, June 2022**

Deployed AI systems often do not work. They can be constructed haphazardly, deployed indiscriminately, and promoted deceptively. However, despite this reality, scholars, the press, and policymakers pay too little attention to functionality. This leads to technical and policy solutions focused on “ethical” or value-aligned deployments, often skipping over the prior question of whether a given system functions, or provides any benefits at all. To describe the harms of various types of functionality failures, we analyze a set of case studies to create a taxonomy of known AI functionality issues. We then point to policy and organizational responses that are often overlooked and become more readily available once functionality is drawn into focus. We argue that functionality is a meaningful AI policy challenge, operating as a necessary first step towards protecting affected communities from algorithmic harm.

CCS Concepts: • **Computing methodologies** → **Machine learning**; • **Applied computing** → **Law, social and behavioral sciences**;

ACM Reference Format:
Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst. 2022. The Fallacy of AI Functionality. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA 14 Pages. https://doi.org/10.1145/3531146.3533158

This fallacy is amplified by opacity



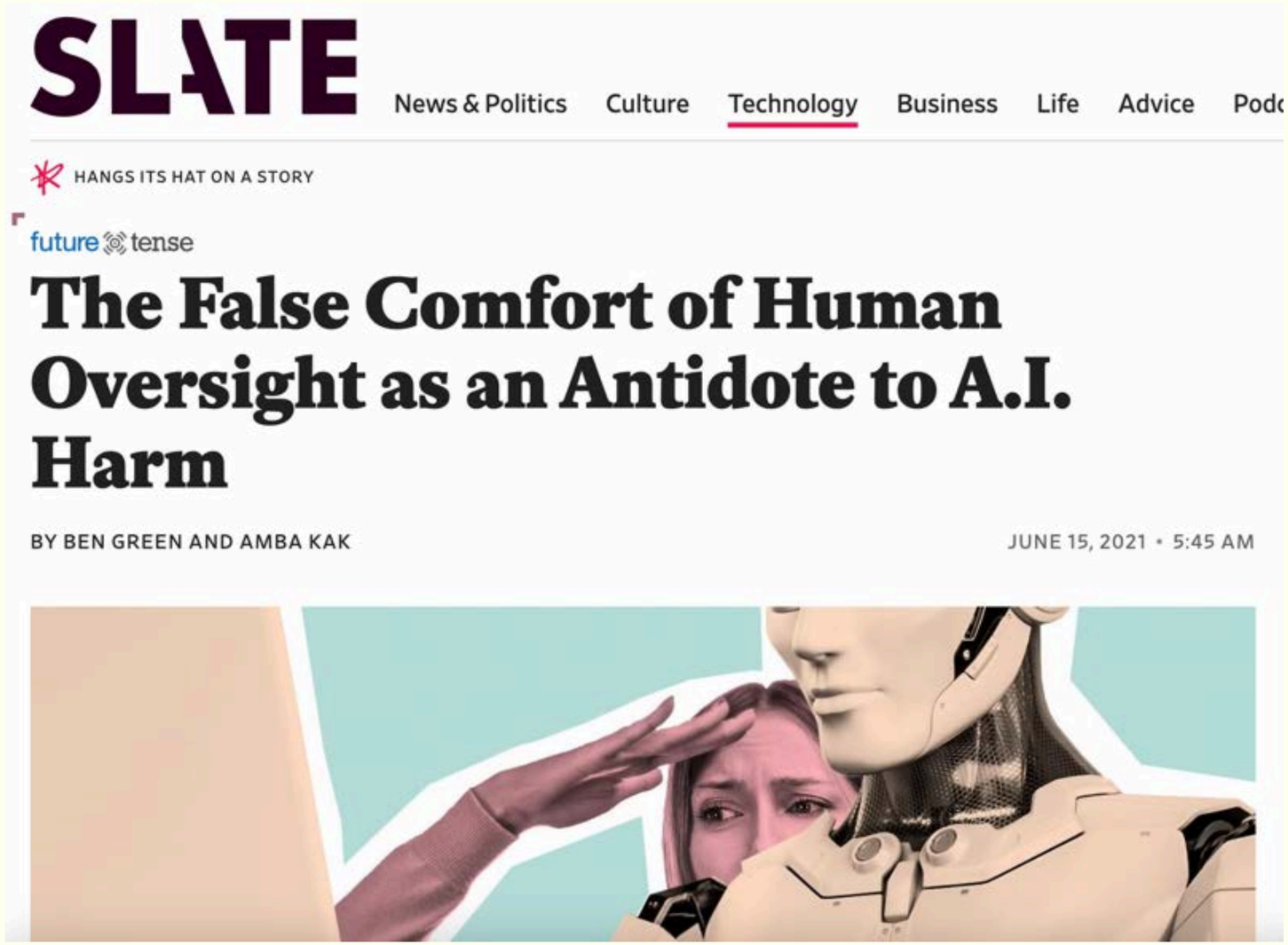
La aplicación del bono social del Gobierno niega la ayuda a personas que tienen derecho a ella

Responde que no cumplen los requisitos a personas que sí los cumplen, y lo hace sin dar explicaciones concretas para resolverlo y pese a haberlo comprobado.

EVA BELMONTE 16 mayo 2019

Technology can make it harder to address errors and performance issues, because of the **assumption** that the system works and the impossibility to prove otherwise due to opacity.

Human oversight as a “false comfort”



automation bias

rubber stamping

disparate interactions

Humans as “moral crumple zones”

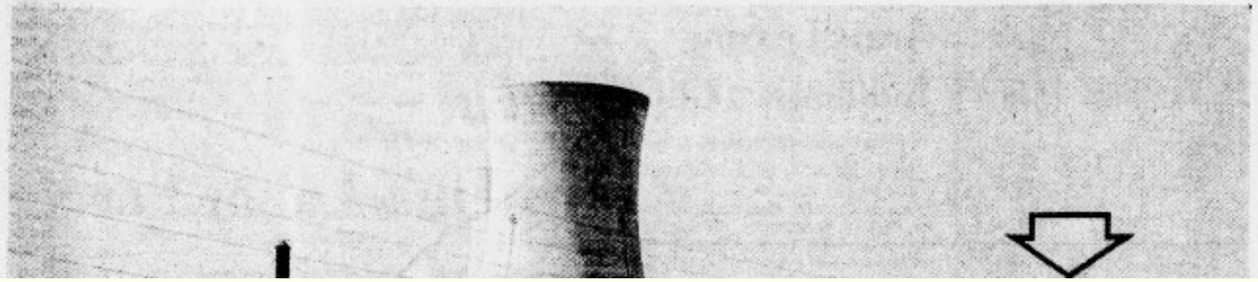
Intelligencer Journal.

Metropolitan Lancaster - 1975 Estimate - U.S. Census 341,300

184th YEAR.—NO. 243 CITY EDITION LANCASTER, PA., THURSDAY MORNING, MARCH 29, 1979. Price 15c — Daily Home Delivered 90c A Week

Worst Leak on Record; Public Not in Danger

Nuclear Mishap at Three Mile Island Spills Radiation Over 16-Mile Area



By CHARLES SHAW
Intelligencer Journal Staff

An accident at the Three Mile Island Nuclear Generating Station early Wednesday morning caused what one government official called "probably the biggest radiation leak" ever from a commercial nuclear plant.

No one was reported seriously injured by the radiation escape into the atmosphere from Unit No. 2 at the station. No residents near the plant, located on the Susquehanna River just north of the Lancaster County boundary in Dauphin County, were evacuated.

The accident that triggered the radiation leak was apparently due to a failure in the plant's cooling system. But just what happened and why, was not completely clear even by late Wednesday.

Humans as “moral crumple zones”

Engaging Science, Technology, and Society 5 (2019), 40-60 DOI:10.17351/ests2019.260

Moral Crumple Zones:
Cautionary Tales in Human-Robot Interaction

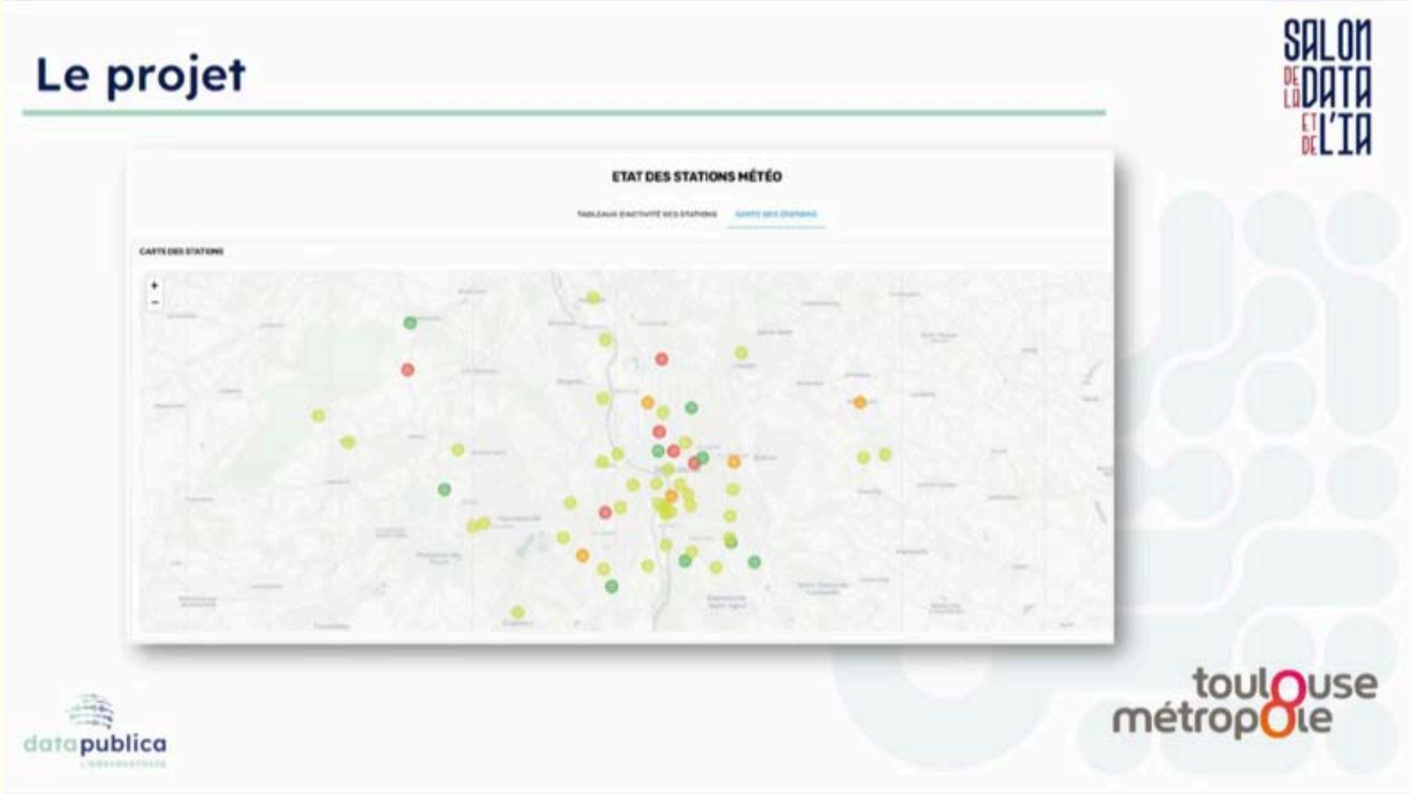
MADELEINE CLARE ELISH
DATA & SOCIETY RESEARCH INSTITUTE

“While the crumple zone in a car is meant to protect the human driver, the moral crumple zone protects the integrity of the technological system, at the expense of the nearest human operator.”

Human oversight is
not a silver bullet.
(including for generative AI!)

Automation may **transform labor**
(not just “make processes more efficient”).

Systems don’t necessarily work as intended: **detecting heat islands in Toulouse**



Sometimes, the critical point of failure is **not technology**.

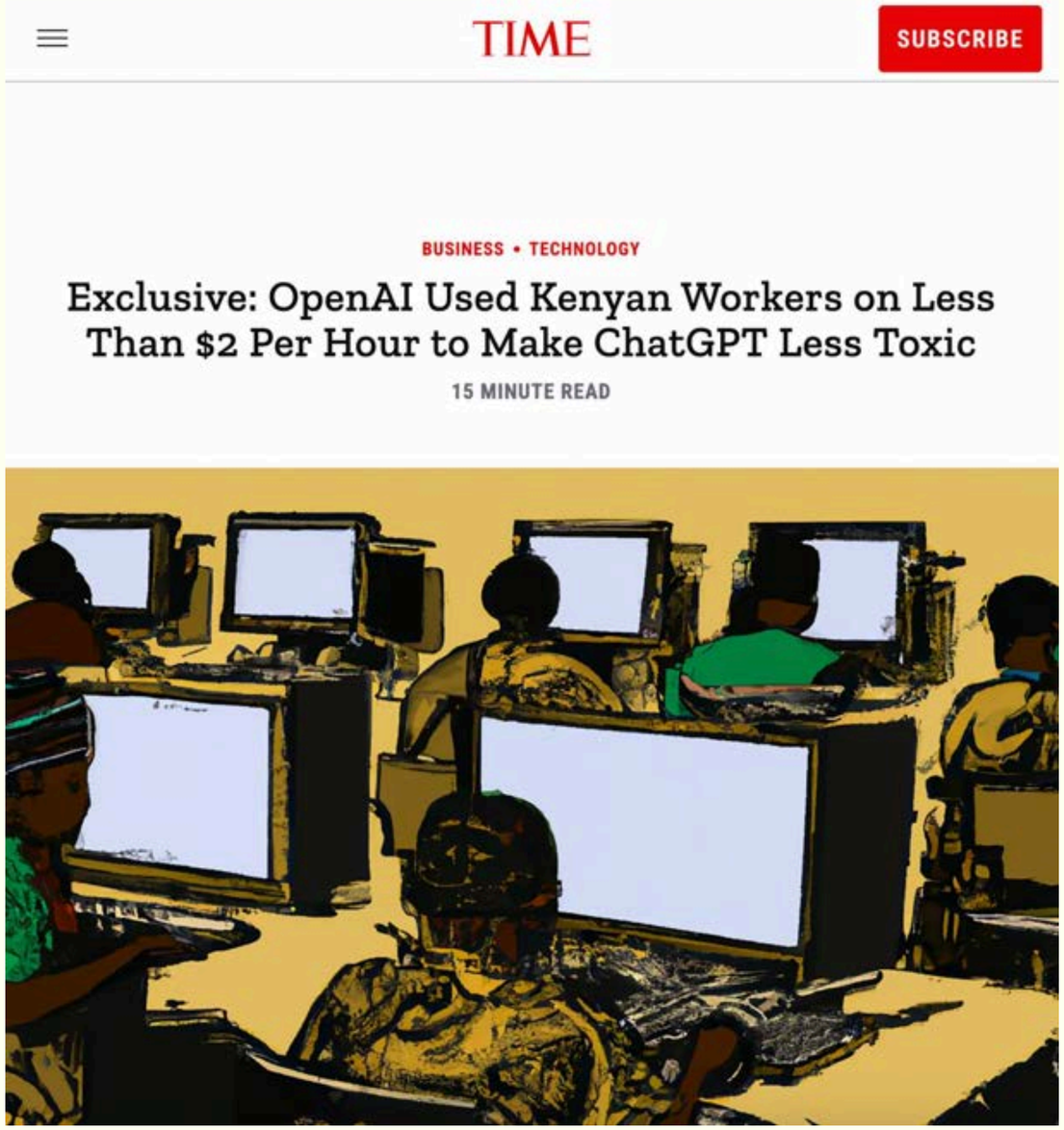
Thinking about systemic implications

Data collection has consequences

But other weights are, at best, inconclusive measures of people's economic standing, and potentially exclusionary. For example, the algorithm assumes that households with higher electricity and water consumption are less vulnerable than households with lower consumption, all else being equal. In NAF's view, higher rates of consumption can imply higher household income. But a family's electricity consumption may be higher because they are *less* well off: for example, a 2020 study of housing sustainability in Amman found that almost 75 percent of low-to-middle income households surveyed lived in apartments with poor thermal insulation, making them more expensive to heat.[141]

Low-income families may also use older, more energy-intensive appliances because they cannot afford to replace them. The government's own social protection strategy indicates that less than 4 percent of people in the poorest decile can afford "cost-saving assets" such as solar water heaters.[142] Nearly everyone Human Rights Watch interviewed about their electricity usage indicated that they were using 500 – 600 kWh per month – well above the median usage of 300 kWh per month.[143]

Global inequalities



TECH CARTOGRAPHIES

YOUR CLOUD IS IN TERRITORIES

The Internet is nothing like a cloud. It is a physical structure, geolocated and crossed by power relations. Who holds the power in this structure? Who has access to this technology? Who profits, who loses, who consumes, who regulates? The Internet is also a territory in dispute, a clash that affects our offline struggles.

FOR ALGORITHMS TO WORK,

all of us, technology consumers worldwide, are doing invisble labour for these companies through...

clicks · comments · shares · likes · contacts · posts

Hate speech and disinformation boost engagement

SURVEILLANCE CAPITALISM

psycho-social problems · spread of hate speech · surveillance · disinformation · manipulation of personal information

DIGITAL COLONIALISM

BIG TECH PROFITS

DEMO CRACY

CODING RIGHTS

CANADA

K Cs

USA

Meta · Microsoft · amazon · Apple · Google · IBM · DELL · hp

None of the CEOs of the big tech companies is a woman and, with rare exceptions, they are all male, white, cisgender, heterosexual, capitalist and capacitist. These are the worldviews transposed to software and algorithms, which are something like the soul of the electronic devices we use. These technologies operate on the logics of data extraction from bodies and territories and violently erase diversities.

One of the largest Google data centers in the world is located in The Dalles, Oregon, USA. Its water consumption has nearly tripled in the past 5 years, using over 25% of the water supplied to the city. The consumption is expected to increase as there are plans for two more data centers in the coming years.

There are over 420 submarine cables deployed around the globe. Their distribution follows the routes of telegraphs and colonial navigations. The largest cable owner is the American company AT&T, followed by China Telecom. However, in recent years, 80% of the investment in new cables comes from Facebook and Google.

DATA CENTERS ACCOUNT FOR ABOUT 4% OF GLOBAL ENERGY CONSUMPTION AND 1% OF GLOBAL GREENHOUSE GAS EMISSIONS.

MEXICO Li · **CUBA** · **PANAMA** Cu Au · **COLOMBIA** Au · **VENEZUELA** Cu Ta Al Au · **ECUADOR** Sn Cu · **PERU** · **BOLIVIA** Li · **CHILE** · **BRAZIL** Au Ta · **PARAGUAY**

Gold is illegally extracted on Kayapó, Munduruku and Yanomami indigenous lands, now traded by refiners supplying Apple, Microsoft, Google and Amazon.

IRELAND · **UK** · **EUROPE**

RUSSIA K

MORE THAN HALF OF THE WORLD'S ELECTRONICS MANUFACTURING IS IN CHINA.

Virtually all circuit boards are printed there. Even the United States is dependent on the production of basic components in China.

TURKEY · **TUNISIA** · **LEBANON** · **PALESTINE** · **ISRAEL** · **IRAN**

surveillance softwares

MOROCCO · **SIERRA LEONE** Au C · **MALI** · **ALGERIA** Au · **CÔTE D'IVOIRE** · **GHANA** · **NIGERIA** · **EGYPT** · **SUDAN** · **SAUDI ARABIA** · **PAKISTAN**

D.R. CONGO W Co · **RWANDA** · **KENYA** · **ETHIOPIA** Au Ta · **ANGOLA** C · **ZAMBIA** Cu · **TANZANIA** Au · **SOUTH AFRICA** Au Pt C

INDIA · **CHINA** · **THAILAND** · **MALAYSIA** · **PHILIPPINES**

lenovo · Baidu 百度 · ByteDance · HUAWEI · WeChat

SOUTH KOREA SAMSUNG · **JAPAN** NEC TOSHIBA SONY EPSON · **TAIWAN** acer

China, major electronics manufacturer for the world, is the largest producer of e-waste. In second place comes the USA. But unlike China, which is also the largest importer of such waste, the US exports its toxic waste to India, China, and several countries in Africa.

Many content moderation operations are outsourced to countries such as India, Kenya, the Philippines, and other Southeast Asian nations, where there is a large pool of English-speaking workers. The work conditions of these "cleaners" are abusive, marked by low pay, secrecy, and mental health consequences. But despite NDAs and alleged union-busting attempts, workers are unionizing and taking Big Tech to court.

Au Cu · Li Al

2034

2036

HOUSEHOLDS WITH

codingrights.org

What now?

Make conscious choices: what are your success metrics? fairness measures? what infrastructure are you choosing?

Make conscious choices: what are your success metrics? fairness measures? what infrastructure are you choosing?

Be transparent about those choices

Make conscious choices: what are your success metrics? fairness measures? what infrastructure are you choosing?

Be transparent about those choices (an EU AI Act requirement!)

Make conscious choices: what are your success metrics? fairness measures? what infrastructure are you choosing?

Be transparent about those choices (an EU AI Act requirement!)

Evaluate **in context**

Make conscious choices: what are your success metrics? fairness measures? what infrastructure are you choosing?

Be transparent about those choices (an EU AI Act requirement!)

Evaluate **in context**

Open up your imagination in problem solving

Thank you!

SOIZIC.PENICAUD@SCIENCESPO.FR
CONTACT@ODAP.FR