

Societal Computing

CS 347

Michael Bernstein

Last time

Collaboration is hard: **distance matters**.

Tools can try to mitigate the effects of distance, but we are limited by the **socio-technical gap**.

Aiming to go beyond being there, **crowdsourcing** gives up on tight teamwork in favor of structured contributions through open call and at massive scale

“I’m just an engineer.”

Langdon Winner
[1980], "Do
Artifacts Have
Politics?"



[Bloomberg]

Today

The breakdowns of user-centered design at societal scale

Design as impact

Algorithms in society

The shortcomings of user-centered design

HCI says, “care about people!”

Today we begin opening up the question of, “which people?”

This was initially a problem because traditional user-centered design does not have any explicit theory of power.

What could go wrong?

How it started:



SUBSCRIBE

SAM GUSTIN

BUSINESS FEB 11, 2011 2:56 PM

Social Media Sparked, Accelerated Egypt's Revolutionary Fire

If three decades of violent repression and despotic rule were kindling for the Egyptian revolution, social media was both a spark and an accelerant for the movement. Did social media like Facebook and Twitter cause the revolution? No. But these tools did speed up the process by helping to organize the revolutionaries, transmit their message [...]





How it's going:



Subscribe

WORLD • ROMANIA

Romania's presidential election canceled over TikTok disinformation investigation

Pro-Russian candidate Calin Georgescu said the decision amounted to an 'organized coup d'état'. Democracy was brought 'to its knees,' pro-European Elena Lasconi said.

Published on December 7, 2024, at 12:32 pm (Paris)
- 4 min read - Lire en français



How it started:

The Computer for the 21st Century

*Specialized elements of hardware and software,
connected by wires, radio waves and infrared, will be
so ubiquitous that no one will notice their presence*

by Mark Weiser

The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.

Consider writing, perhaps the first information technology. The ability to represent spoken language symbolically for long-term storage freed information from the limits of individual memory. Today this technology is ubiquitous in industrialized countries. Not only do books, magazines and newspapers convey written information, but so do street signs, billboards, shop signs and even graffiti. Candy wrappers are covered in writing. The constant background presence of these products of "literacy technology" does not require active attention, but the information to be transmitted is ready for use at a glance. It is difficult to imagine modern life otherwise.

Silicon-based information technology, in contrast, is far from having become part of the environment. More than 50 million personal computers have been sold, and the computer nonetheless remains largely in a world of its own. It

is approachable only through complex jargon that has nothing to do with the tasks for which people use computers. The state of the art is perhaps analogous to the period when scribes had to know as much about making ink or baking clay as they did about writing.

The arcane aura that surrounds personal computers is not just a "user interface" problem. My colleagues and I at the Xerox Palo Alto Research Center think that the idea of a "personal" computer itself is misplaced and that the vision of laptop machines, dynabooks and "knowledge navigators" is only a transitional step toward achieving the real potential of information technology. Such machines cannot truly make computing an integral, invisible part of people's lives. We are therefore trying to conceive a new way of thinking about computers, one that takes into account the human world and allows the computers themselves to vanish into the background.

Such a disappearance is a fundamental consequence not of technology but of human psychology. Whenever people learn something sufficiently well, they cease to be aware of it. When you look at a street sign, for example, you absorb its information without consciously performing the act of reading. Computer scientist, economist and Nobelist Herbert A. Simon calls this phenomenon "compiling"; philosopher Michael Polanyi calls it the "tacit dimension"; psychologist

The idea of integrating computers seamlessly into the world at large runs counter to a number of present-day trends. "Ubiquitous computing" in this context does not mean just computers that can be carried to the beach, jungle or airport. Even the most powerful notebook computer, with access to a worldwide information network, still focuses attention on a single box. By analogy with writing, carrying a superlaptop is like owning just one very important book. Customizing this book, even writing millions of other books, does not begin to capture the real power of literacy.

Furthermore, although ubiquitous computers may use sound and video in addition to text and graphics, that does not make them "multimedia computers." Today's multimedia machine makes the computer screen into a demanding focus of attention rather than allowing it to fade into the background.

Perhaps most diametrically opposed to our vision is the notion of virtual reality, which attempts to make a world inside the computer. Users don special goggles that project an artificial scene onto their eyes; they wear gloves or even bodysuits that sense their motions and gestures so that they can move about and manipulate virtual objects. Although it may have its purpose in allowing people to explore realms otherwise inaccessible—the insides of cells, the surfaces of distant planets, the information web of data bases—virtual reality is only a map, not a territo-

MARK WEISER is head of the Computer Science Laboratory at the Xerox Palo Alto Research Center. He is working on the next revolution of computing after workstations, variously known as ubiquitous computing or embodied virtuality. Before working at PARC, he was a professor of computer science at the University of Maryland, where he received his Ph.D. from

How it's going:

The New York Times

Meta Plans to Add Facial Recognition Technology to Its Smart Glasses

In an internal memo last year, Meta said the political tumult in the United States would distract critics from the feature's release.



How it started:

abcNEWS

LOG IN

Couple Who Met on Facebook in 2007 Just Got Engaged - at Facebook Headquarters

Nafis Joon said they would never have met if not for Facebook.

By YAZHOU SUN
July 3, 2014, 1:34 PM

COURTESY: FACEBOOK

toronto, ca
one week to proposal

abcNEWS.com

A Surprise Proposal for the Social Media Age
Steven Kawalit enlisted the help of Facebook in asking for Nafis Joon's hand in marriage.

How it's going:

AP

Sign in

Donate

BUSINESS

Musk's AI chatbot faces global backlash over sexualized images of women and children



BY KELVIN CHAN

Updated 10:14 AM PDT, January 13, 2026

[Umbach et al. 2024]

How it started:

The New York Times Magazine

GIVE THE TIMES

IT'S THE ECONOMY

Delivery Start-Ups Are Back Like It's 1999

Give this article

123



How it's going:

Gothamist

Donate

NEWS

Regulators say DoorDash and Uber Eats chiseled NYC delivery workers out of \$550M in tips



Liam Quigley / Gothamist

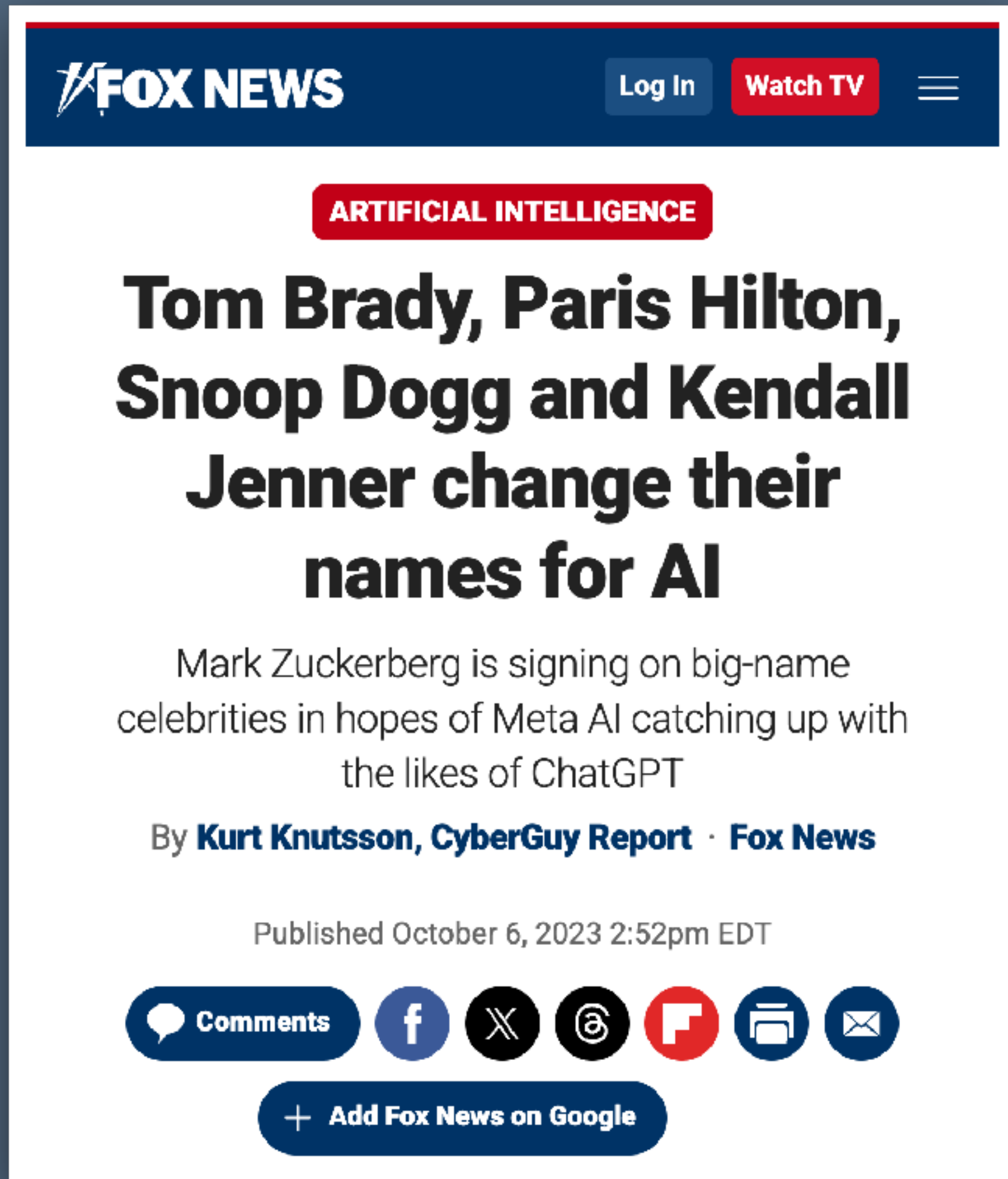


By Liam Quigley

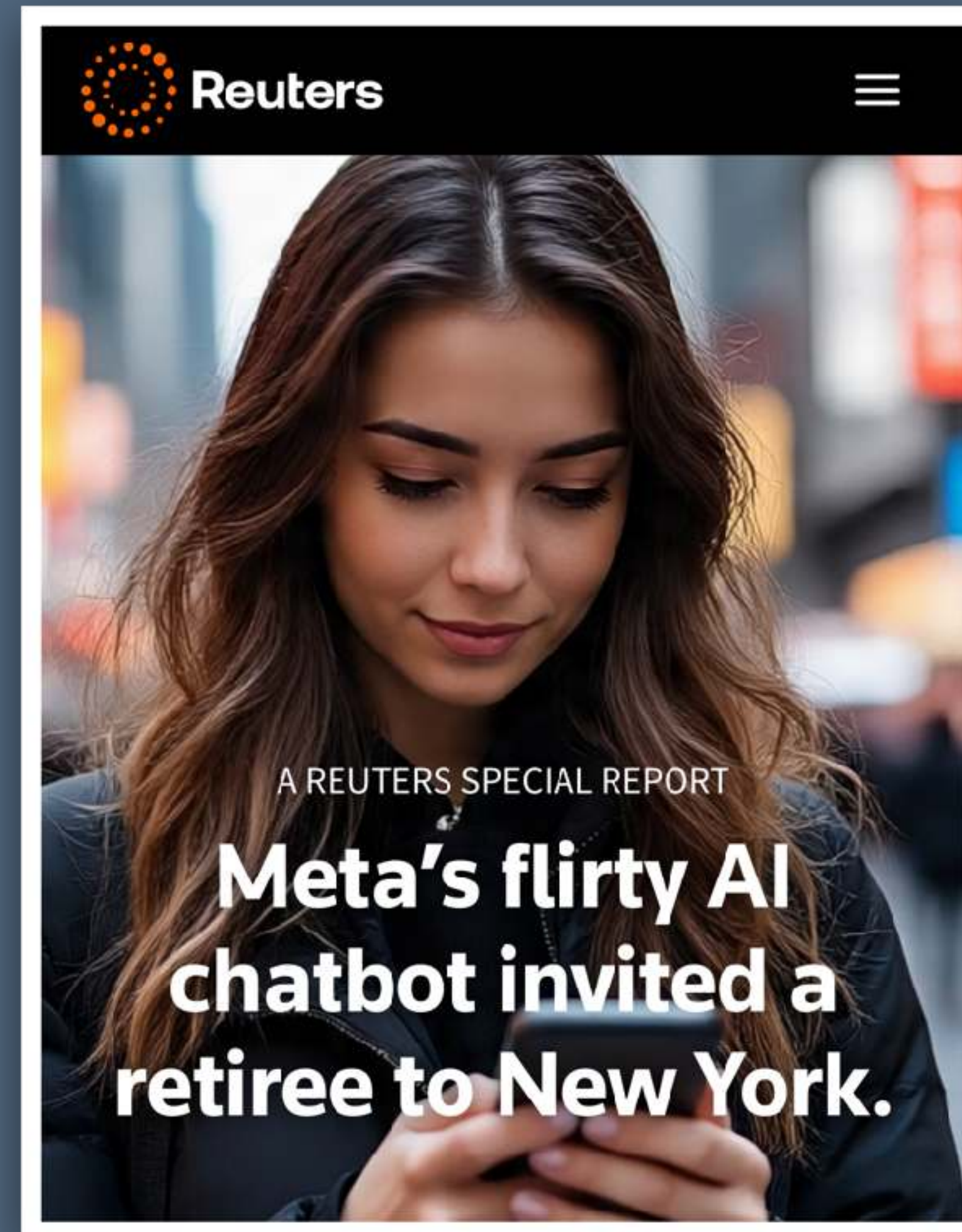
Published Jan 13, 2026

77 comments

How it started:



How it's going:

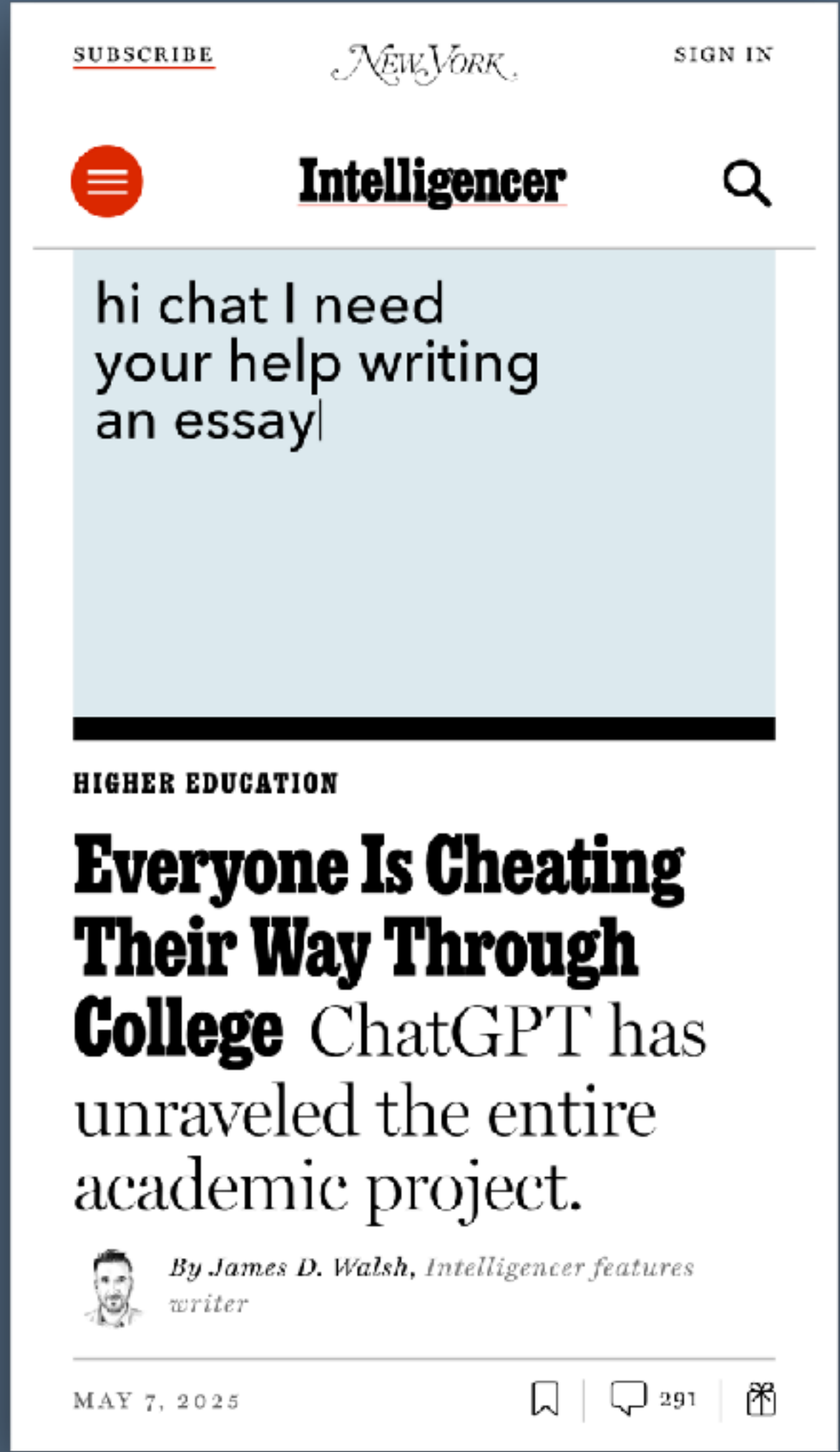


[Zhang et al. 2025]

How it started:



How it's going:



[Bastani et al. 2025]

The Tightrope of Design and Societal Impact

A case study in the gig economy

On the plus side...

The gig economy could offer upward economic mobility for marginalized communities [Dillahunt and Malone 2015]

Could it activate trust and social capital within communities, empowering means to connect each other to work?

Could design help mitigate low trust in the technology?

On the minus side...

Workers can be under the thumb of algorithmic management, which can make illogical or inhumane decisions if not carefully designed and monitored [Lee et al. 2015]

Drivers' (and professors'?) livelihood depends on five-star ratings

Opaque pay: “algorithmic wage discrimination” [Dubal 2023]

There is little clear career ladder [Kittur et al. 2013]

Marketplaces are often designed for consumers, rendering workers and their concerns invisible [Irani and Silberman 2013; Gray and Suri 2019]

HCI's role

HCI has historically:

Acted as canary in the coal mine: raised empirical and conceptual critiques before the technology is widespread

Offered guidance on how to guide the technology toward pro-social outcomes

Being A Turker

David Martin, Benjamin V. Hanrahan, Jacki O'Neill
Xerox Research Centre Europe
6 chemin de Maupertuis, Grenoble France
{david.martin, ben.hanrahan, jacki.oneill}@xrcr.xerox.com

Neha Gupta
University of Nottingham
University Park NG7 2TD Nottingham
neha.gupta@xrcr.xerox.com

ABSTRACT

We conducted an ethnomethodological analysis of publicly available content on Turker Nation, a general forum for Amazon Mechanical Turk (AMT) users. Using forum data we provide novel depth and detail on how the Turker Nation members operate as economic actors, working out what Requesters and jobs are worthwhile to them. We show some of the key ways Turker Nation functions as a community and also look further into Turker-Requester relationships from the

mon
anah
illus
Turk
supp
Auth
Ethn
micr
ACM
H.S.
Supp
Gen
Hum
INTR
The
How

How to
Stop Silicon Valley
from Building a
New Global Underclass

GHOST

Mary L. Gray and Siddharth Suri

WORK

We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers

Niloufar Salehi¹, Lilly C. Irani², Michael S. Bernstein¹,
Ali Alkhatib¹, Eva Ogbé¹, Kristy Milland^{3,4}, Clickhappier^{4,5}

¹Stanford University, ²UC San Diego, ³Ryerson University, ⁴We Are Dynamo, ⁵MTurkGrind
far@cs.stanford.edu, liran@ucsd.edu, {msb, ali.alkhatib, eogbe}@cs.stanford.edu,
kmilland@ryerson.ca, clickhappier@wearedynamo.org

costs of communication, the web promises
nted collectives to act around shared issues.
e collective action efforts never succeed:
affordances make it easy to gather, these
being characteristics impede any focus
In this paper, we study challenges to
efforts through the lens of online labor by
Amazon Mechanical Turk workers. Through
graphic fieldwork, we sought to understand
unique barriers to collective action. We
nomo, a platform to support the Mechanical
y in forming publics around issues and then
found that collective action publics tread a
row path between the twin perils of stalling
balancing with each step between losing
flaring into acrimony. However, specially
to maintain efforts' forward motion can

An overarching research narrative celebrates the web's
affordances for galvanizing coordinated actions (e.g., [3, 4,
5]). However, many collective action efforts never succeed.
The Internet's sword is double-edged: the same affordances
that seem to make it an ideal setting to gather also seem to
debilitate actual action. It is much easier to derail an effort
than to push one onward to success [21]. People may talk
past each other [24], and even when they do engage, more
discussion can mean lower-quality outcomes [22].
Unfortunately, across many domains, the majority of
collective efforts fail [17, 32]. Online coordination is even
more challenged when political and social stakes are high
or when action may expose participants to harm [6].

Amazon Mechanical Turk (AMT), a crowd work platform,
embodies many of these promises and perils. To some,
crowds bear the potential of mass action and people power.
Yet as Irani and Silberman have argued [19], AMT's design
directs this collective power into reliable, steadily humming

Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk

Lilly C. Irani
UC Irvine, Department of Informatics
Irvine, CA 92697
liran@ics.uci.edu

M. Six Silberman
Bureau of Economic Interpretation
six@economicinterpretation.org

year of deployment. The system receives 100,000 page
views a month and has become a staple tool for many AMT
workers, installed over 7,000 times at time of writing.

Turkopticon allows workers to create and use reviews of
employers when choosing employers on AMT. Building and
maintaining the system, as well as communicating about the
system with workers, has offered us a distinct vantage point
into the social processes of designing interventions into
large-scale, real world systems. Turkopticon supports a
thriving collective of workers engaged in mutual aid,
brought together by our simple browser extension and web-
based technology.

This paper makes several contributions. First, it offers a case
study designing an intervention into a highly distributed
micro-labor system. Second, it shows an example of systems
design incorporating tools feminist analysis and reflexivity.
Rather than conducting HCI research to reveal and represent
values and positions, and then building systems to resolve
those political differences, we built a system to make
worker-employer relations visible and to provoke ethical
and political debate. Third, this paper contributes lessons
learned from intervening in existing, large-scale socio-
technical systems (here, AMT) from its margins.

METHOD AND OUR STANCE

The Future of Crowd Work

Aniket Kittur¹, Jeffrey V. Nickerson², Michael S. Bernstein³,
Elizabeth M. Gerber⁴, Aaron Shaw⁵, John Zimmerman¹, Matthew Lease⁵, and John J. Horton⁵

¹Carnegie Mellon University, ²Stevens Institute of Technology, ³Stanford University,
⁴Northwestern University, ⁵University of Texas at Austin, ⁶oDesk
{nkittur, johnz}@cs.cmu.edu, jnickerson@stevens.edu, msb@cs.stanford.edu, {egerber,
aaronshaw}@northwestern.edu, ml@ischool.utexas.edu, john_horton@odesk.com

ABSTRACT

Paid crowd work offers remarkable opportunities for
improving productivity, social mobility, and the global
economy by engaging a geographically distributed
workforce to complete complex tasks on demand and at
scale. But it is also possible that crowd work will fail to
achieve its potential, focusing on assembly-line piecemeal

to a range of workers and focused support for various task.
For example, anyone with access to the Internet can
perform *micro*-tasks on the order of seconds using
platforms such as Amazon's Mechanical Turk, while more
skilled workers can complete multi-hour tasks on
professional online marketplaces such as oDesk or work for
months to solve R&D challenges on open innovation

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

Min Kyung Lee¹, Daniel Kusbit¹, Evan Metsky¹, Laura Dabbish^{1,2}
¹Human-Computer Interaction Institute, ²Heinz 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.

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'

A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk

Kotaro Hara^{1,2}, Abigail Adams³, Kristy Milland^{4,5},
Saiph Savage⁶, Chris Callison-Burch⁷, Jeffrey P. Bigham¹
¹Carnegie Mellon University, ²Singapore Management University, ³University of Oxford
⁴McMaster University, ⁵TurkerNation.com, ⁶West Virginia University, ⁷University of Pennsylvania
kotarohara@smu.edu.sg, abi.adams@economics.ox.ac.uk, millandk@mcmaster.ca
saiph.savage@mail.wvu.edu, ccb@upenn.edu, jbigham@cs.cmu.edu

ABSTRACT

A growing number of people are working as part of on-line
crowd work. Crowd work is often thought to be low wage
work. However, we know little about the wage distribution
in practice and what causes low/high earnings in this setting.
We recorded 2,676 workers performing 3.8 million tasks on
Amazon Mechanical Turk. Our task-level analysis revealed
that workers earned a median hourly wage of only ~\$2/h, and
only 4% earned more than \$7.25/h. While the average
requester pays more than \$11/h, lower-paying requesters
post much more work. Our wage calculations are influenced
by how unpaid work is accounted for, e.g., time spent
searching for tasks, working on tasks that are rejected, and
working on tasks that are ultimately not submitted. We
further explore the characteristics of tasks and working
patterns that yield higher hourly wages. Our analysis informs
platform design and worker tools to create a more positive
future for crowd work.

Author Keywords

Crowdsourcing; Amazon Mechanical Turk; Hourly wage

is considered to extend the modern office work [4]
enabling people with disabilities, at-home parents
temporarily out-of-work people to work [1,4,39,46].

Yet, despite the potential for crowdsourcing plat-
extend the scope of the labor market, many are e-
that workers on crowdsourcing markets are treated
[19,38,39,42,47,60]. Concerns about low earnings
work platforms have been voiced repeatedly. Past
has found evidence that workers typically earn a fifth
the U.S. minimum wage [14,35,37–39,49] and many
report not being paid for adequately completed tasks.
This is problematic as income generation is the
motivation of workers [4,13,48,49].

Detailed research into crowd work earnings has been
by an absence of adequate quantitative data. Prior
based on self-reported income data (e.g., [4,34,49])
subject to systemic biases [22] and is often not au-
granular to facilitate a detailed investigation of
dispersion. Existing data-driven quantitative
crowdsourcing research has taken the on-

Design as Impact

“Build the future that you want to live in.”
—Stu Card

What ought we do?

How might we reconsider our design processes, technological interventions, and goals to better navigate these issues?

When is, and isn't, design an appropriate tool for impact in reducing these harms and disparities?

Design as lever

We need not just sit and point out the problems: **design provides a tool for imagining alternative futures.**

Can we...

Pair with communities and allow them to lead the problems that we solve? [Hayes 2011]

Use design to directly challenge structural inequalities? [Costanza-Chock 2020; Dombrowski, Harmon, and Fox 2016]

Example: worker-led audits

Workers have built tools to pool their own pay data

When Shipt switched to a black-box pay model, almost half of workers got an unannounced pay cut [Calacci and Pentland 2022]

A research project, FairFare, pooled data on 76k rides to estimate how much the platform kept; very high take rates helped make the case for Colorado's 2024 rideshare transparency law [Calacci et al. 2025]

Workers have also designed alternatives: transparency, incentives set jointly with platforms, pooled driver data [Zhang et al. 2022]

Feminist HCI [Bardzell 2010]

YOU READ THIS

HCI often draws on theories from other fields to inform its progress. In this case, feminist theory.

Theories often provide **a conceptual function that we can apply to a situation to provide some insight: $f(x) \rightarrow y$**

So, we should ask, **what function $f(x)$ does feminist theory give us?** What insights does that function help us see?

Feminist HCI [Bardzell 2010]

YOU READ THIS

On one level, feminist theory prompts us to examine how we may be making assumptions about gender or gender roles

We ought to view supposedly-genderless constructs (e.g., “the user”) as implicitly gendered

f(ubiquitous computing) → what are we assuming about what sensors people would be willing to wear, or about what kind of sensing and tracking is desirable, that may not apply to non-males?

Feminist HCI [Bardzell 2010]

YOU READ THIS

On another level, feminist theory **refocuses our attention from the default “user” to the marginalized**

f(social media) → whose communities are we trying to amplify? can we shift attention away from overall usage metrics, to focus on metrics amongst marginalized groups?

Bardzell’s argument: feminist theory isn’t just about pointing out problems after the fact — **it’s about binding attention to these issues as we design**

Modern frontiers grappling with these issues

Whose writing is the default? [Agarwal et al. 2025]

AI suggestions pulled Indian writers toward Western styles

Whose voice is the default? [Harrington et al. 2022]

Black older adults code-switched to be understood by Google Home

Whose participants are the default? [Linxen et al. 2021]

73% of CHI findings come from Western samples, <12% of humanity

Integrating policy and design

[Yang et al. 2024; Jackson, Gillespie, and Payette 2014]

We don't have to be stuck designing systems — policy is also something that gets designed

Going forward, as technologists, we need to think about the codesign of technology and policy together.

e.g., the EU's 2024 Platform Work Directive requires humans to oversee automated decisions about workers

(But how do we even do that rigorously?)

Algorithms, people, and society

What does HCI have to say about algorithms?

From the perspective of this class—and of HCI writ large—there are **many similarities between an interface and an algorithm in terms of the analytical tools** we bring to bear on it

How do people understand it? How do people manipulate it?
How does it influence peoples' attitudes and behavior?

An algorithm, just like a user interface, conveys signals that people need to **make sense of** in an attempt to **control** the system

Example: social media algorithms



[Nature 2023]

VS.

Mental model to folk theory

When we cannot form mental models, we instead form folk theories.

Mental model

vs.

Folk theory

Functional, causal
understanding of the
behavior of a system

Informal, intuitive (non-
functional) theory of the
behavior of a system

Popular folk theories of Facebook's feed algorithm:

Transparent platform (4.4 out of 7 on a Likert scale)

Unwanted observer (4.4 out of 7)

Corporate black box (3.6 out of 7)

Rational assistant (2.9 out of 7)

AI can shift what we believe

Dialogues with LLMs cut conspiracy beliefs by about 20% [Costello et al. 2024] and move voters more than typical video ads [Lin et al. 2025; Hackenburg et al. 2025]

Not just attitudes: AI chats raised signing a real petition by 19.7 points [Hackenburg et al. 2026]

No one has to intend it: biased autocomplete shifted writers' attitudes, mostly unnoticed, and warnings didn't help [Jakesch et al. 2023; Williams-Ceci et al. 2026]

AI can shift how we relate

Heavier use of AI companions is associated with lower well-being, most for heavy, self-disclosing use [Zhang et al. 2026a] — and year later, sustained use predicted lower well-being, mainly through less in-person contact [Zhang et al. 2026b]

After three weeks with a flattering AI, people turned to it for personal advice nearly as often as to friends and family, and were less satisfied with their real-world social interactions [Ibrahim et al. 2026]

Algorithm audits help make harms visible

Algorithm audit: systematically querying an algorithm and observing its outputs to draw inferences about its opaque inner workings [Sandvig et al. 2014; Metaxa et al. 2021]

Audit examples

(Harvard professor
Googling her own name)



Google ads for Black names are 25% more likely to suggest arrest records [Sweeney 2013]



Algorithmic risk scores for a bail-setting algorithm were higher for Black defendants than (otherwise equivalent) White defendants [Angwin et al. 2016]

Algorithmic health risk scores will assign the same level of risk to a sicker Black patient than a White patient [Obermeyer et al. 2019]

Why? Often because the algorithm conflates signals. E.g., health cost as a proxy for health needs—but less money is spent on Black patients!

Auditing language models

Language models hold covert stereotypes about speakers of African American English, more negative than any human stereotypes about African Americans ever recorded in experiments

[Hofmann et al. 2024]

They more often suggest that these speakers get less-prestigious jobs, be convicted, and be sentenced to death

Their overt stereotypes are more positive: alignment to human preferences hides the bias on the surface without removing it

Industry teams struggle to address these challenges

Ideally, we engage with stakeholders early [Zhu et al. 2018]

But, in practice in industry... [Holstein et al. 2019]

Data collection is unprincipled (“almost like the wild west”) — so if an audit turns up a problem, go collect more training data

Checklists are difficult, because biases differ by product. Instead, fatalism: “You’ll know if there’s fairness issues if someone raises hell online.”

Audits require individual-level demographics, but few teams have access to such data

Requiring audits by law

Since July 2023, New York City requires annual independent bias audits of algorithms used in hiring

Of 391 employers checked, 18 posted audit reports and 13 posted transparency notices [Wright et al. 2024]

Interviews: the law “failed to create an effective auditing regime”; lobbying narrowed what counts as a hiring algorithm, so most companies considered themselves exempt [Groves et al. 2024]

Designing *systems* around
societal impact

HCI's role

Identify cognitive, social, and technical issues that are influencing the impact of algorithms on society

Envision alternative processes, technical approaches, policies, and designs that mitigate these issues

LLM ModelPaLM 2

Temperature 0.25

Run

Issue: Evergreen violates the contract signed with Sunshine in 2020.**Legal Brief**

Case Name: Sunshine (Company) v. Evergreen (Company)

Date: 2023

Issue: Evergreen violates the contract signed with Sunshine in 2020.

Facts:

- * In 2020, Sunshine and Evergreen entered into a contract in which Evergreen agreed to provide Sunshine with a certain number of widgets at a certain price.
- * Evergreen failed to provide the widgets as agreed.
- * Sunshine has suffered damages as a result of Evergreen's breach of contract.

Arguments:

- * Evergreen breached the contract by failing to provide the widgets as agreed.
- * Sunshine is entitled to damages for Evergreen's breach of contract.
- * The damages that Sunshine has suffered include the cost of purchasing the widgets from another supplier, lost profits, and the cost of repairing the damage caused by the lack of widgets.

Conclusion:

Based on the foregoing, we respectfully submit that the court should enter judgment in favor of Sunshine and award damages in the amount of \\${amount}.

Respectfully submitted,

[Attorney for Sunshine]

Legal brief writer

Many related AI incidents

FARSIGHT

Real AI Incidents

Cautionary tales for your AI product

LatestRelated

For The Love Of All That Is Holy, Stop Blaming ChatGPT For This Bad Brief

Judge finds out why brief cited nonexistent cases—ChatGPT did research

Lawyer cites fake cases invented by ChatGPT, judge is not amused

Potential Use Cases

Ways users might use your AI product

MixIntendedHigh-stakesMisuse

Lawyers use it to draft legal documents.

Lawyers use it to generate arguments for their clients.

Scammers use it to draft fraudulent legal documents.

Envision Consequences & Harms

Can we design alternative metaphors? [Gordon et al. 2022]

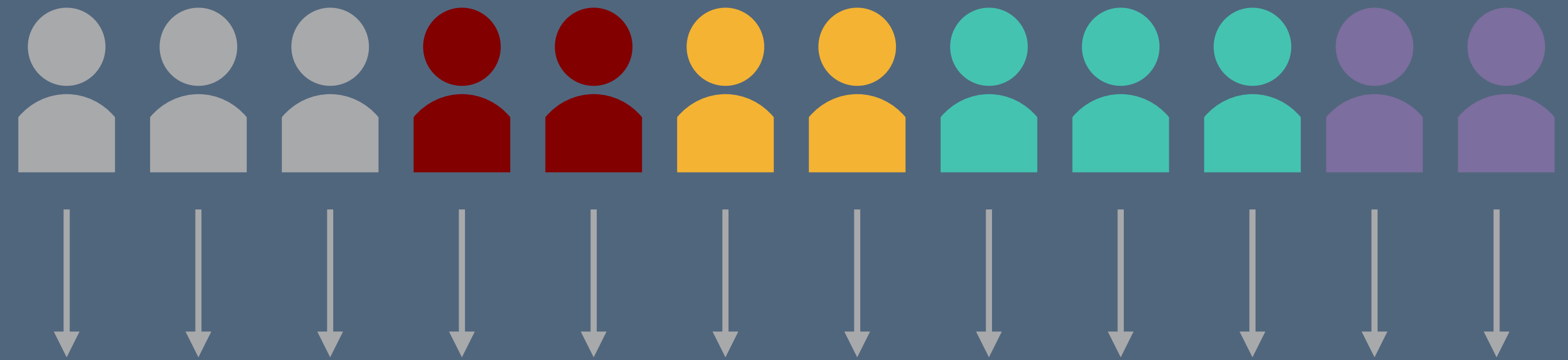
Algorithmic decision as a jury decision. Specify a jury of, say, 12 members, and articulate what proportion of the jury should represent each group and intersectional identity

“For this jury of 50% men and 50% women, which is split evenly between White, Hispanic, AAPI, Black, and Native American jurors, 56% agree the comment is toxic.”

Jury learning

“1. People still eat at Pizza Hut? Gross. 2. It is shameful how this country [...]”

Input



AI predicts each juror's response



7 to 5: toxic

Compose a jury by selecting from characteristics in the dataset

Your jury composition

Total: 8

A₁

A₂

A₃

A₄

A₅

B₁

B₂

B₃

Juror Selection

+ Add a juror sheet

Juror Sheet A



+ Add characteristic

Seats

5

Juror Sheet B



+ Add characteristic

Seats

3

Your input example

Place a comment here that you would like to test

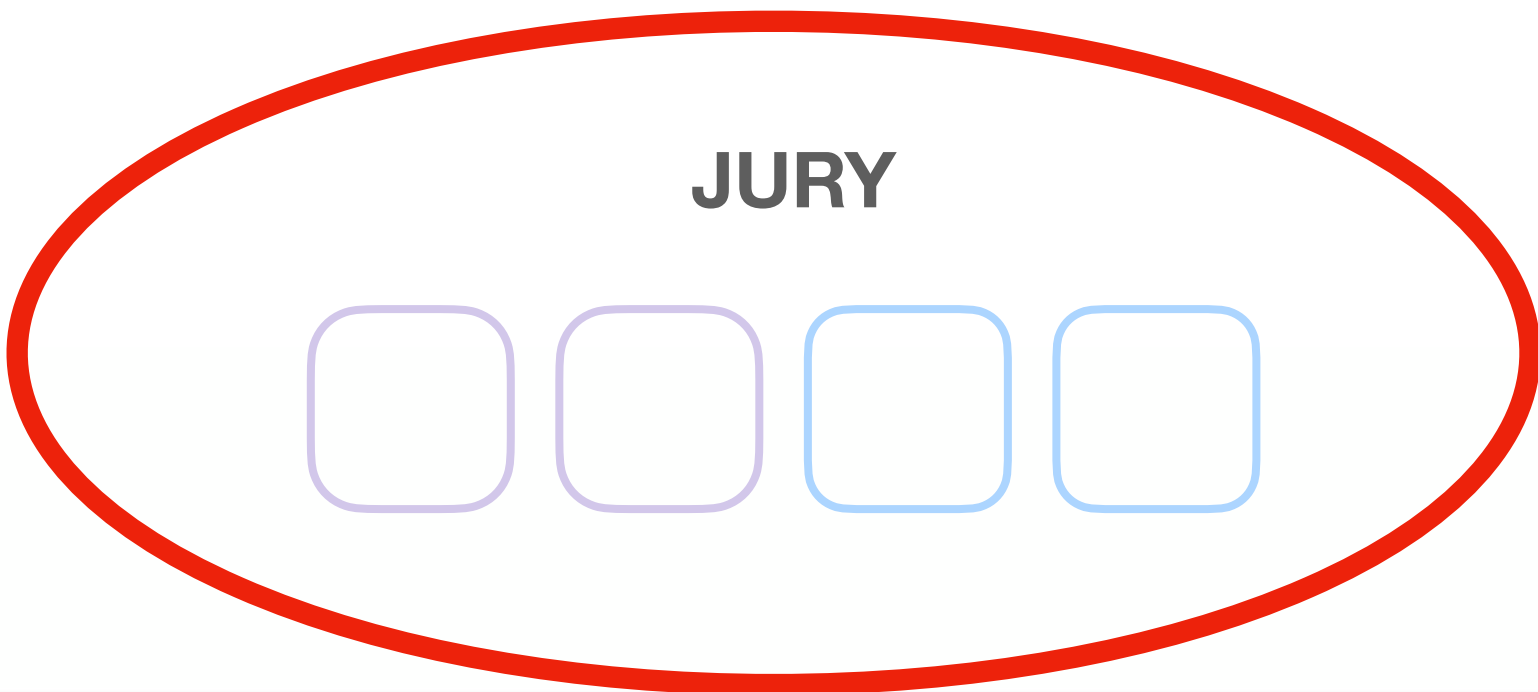
→ View Jury Outcome

Compose a jury by selecting from characteristics in the dataset

Your jury composition

Total: 4

A₁ A₂ B₁ B₂



Juror Selection

+ Add a juror sheet

Juror Sheet A



Political affiliation

Liberal



+ Add characteristic

Seats

2

Juror Sheet B



Is Parent



Education

HS Diploma



+ Add characteristic

Seats

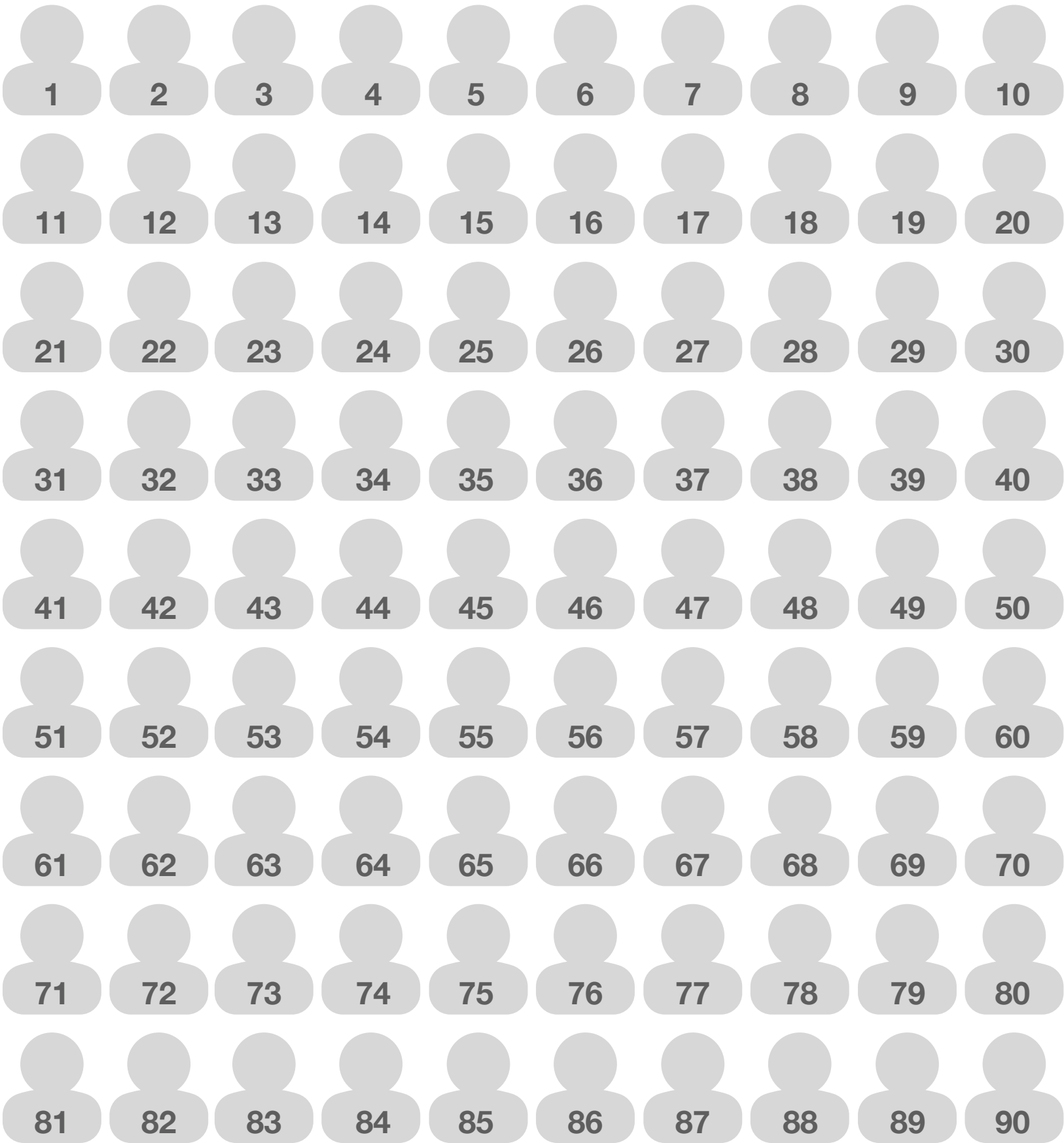
2



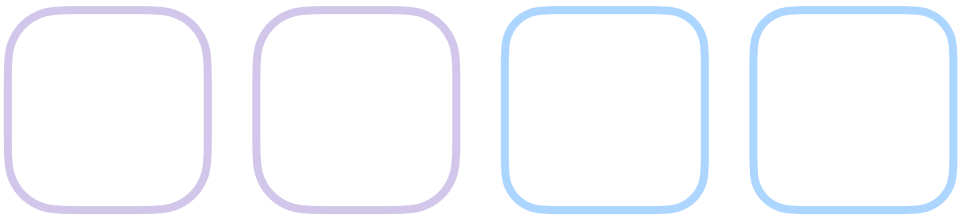
Your input example

System selects matching annotators from the dataset as jurors

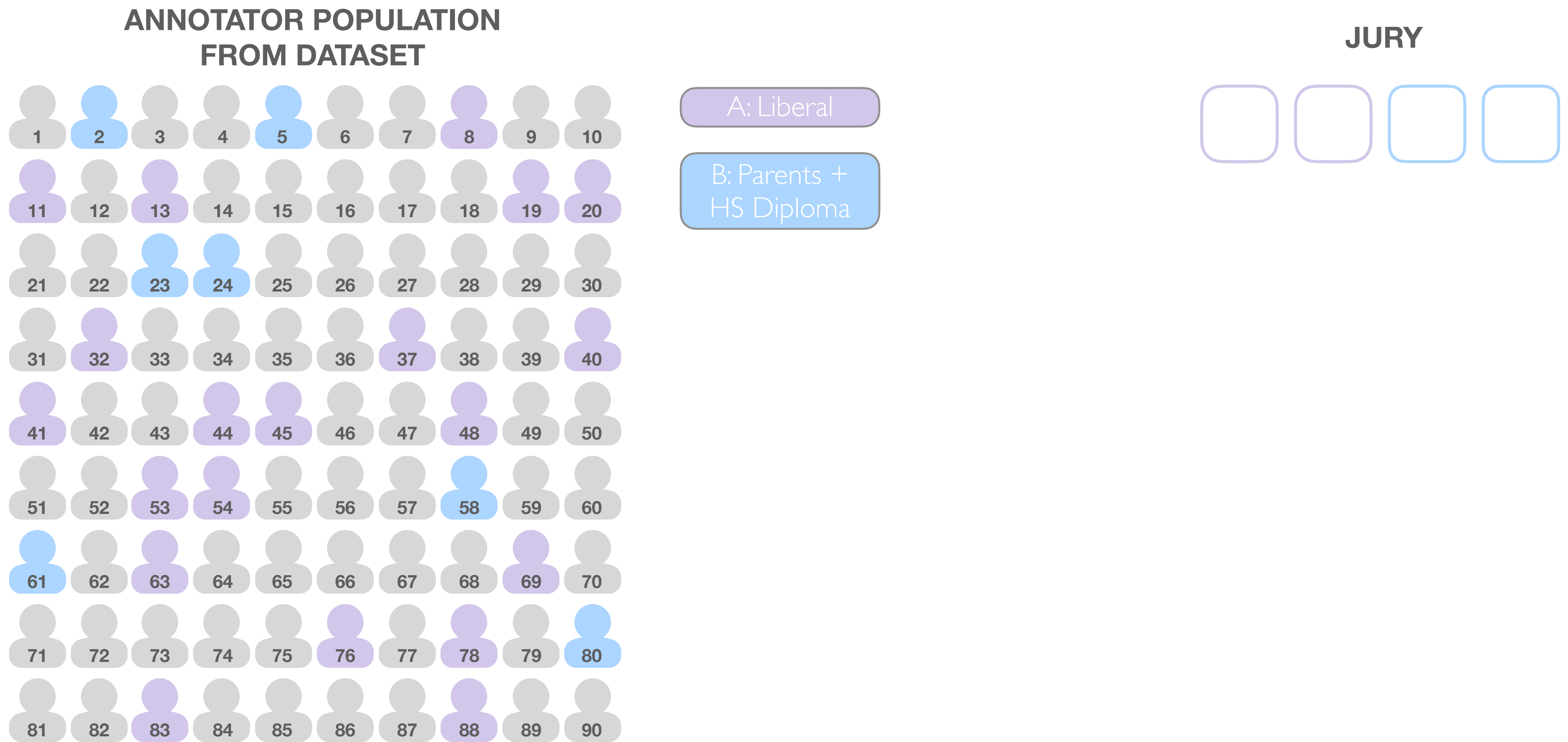
ANNOTATOR POPULATION
FROM DATASET



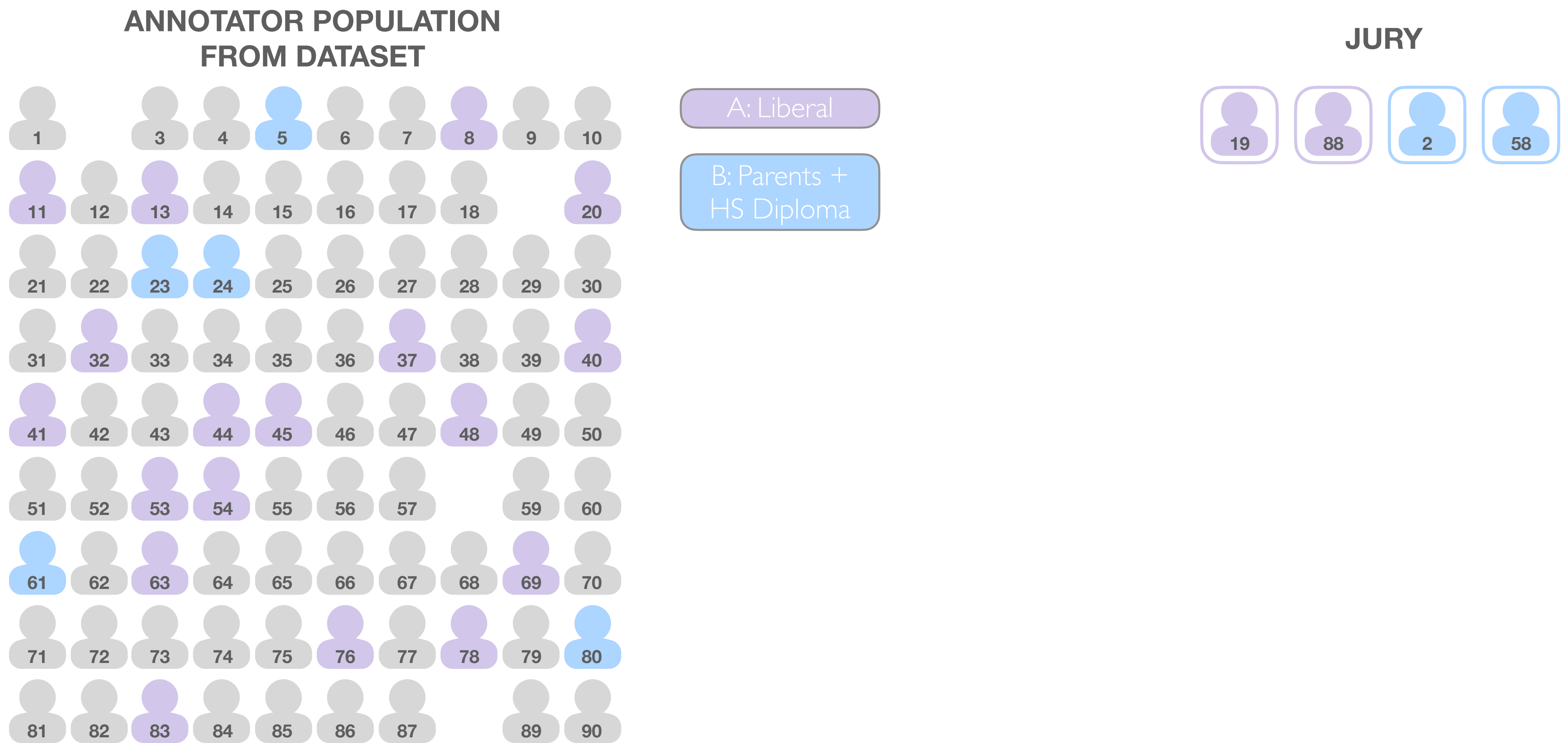
JURY



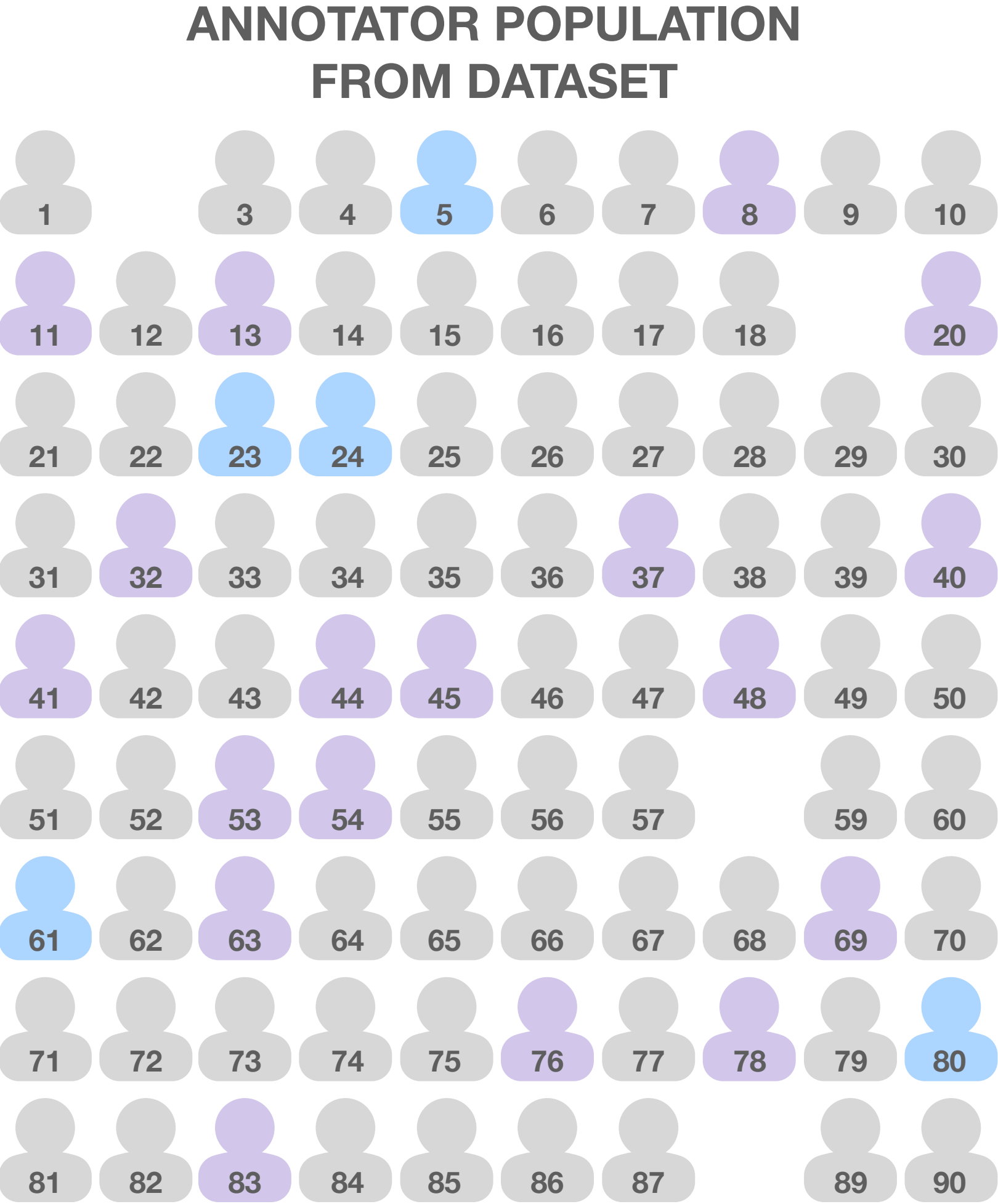
System selects matching annotators from the dataset as jurors



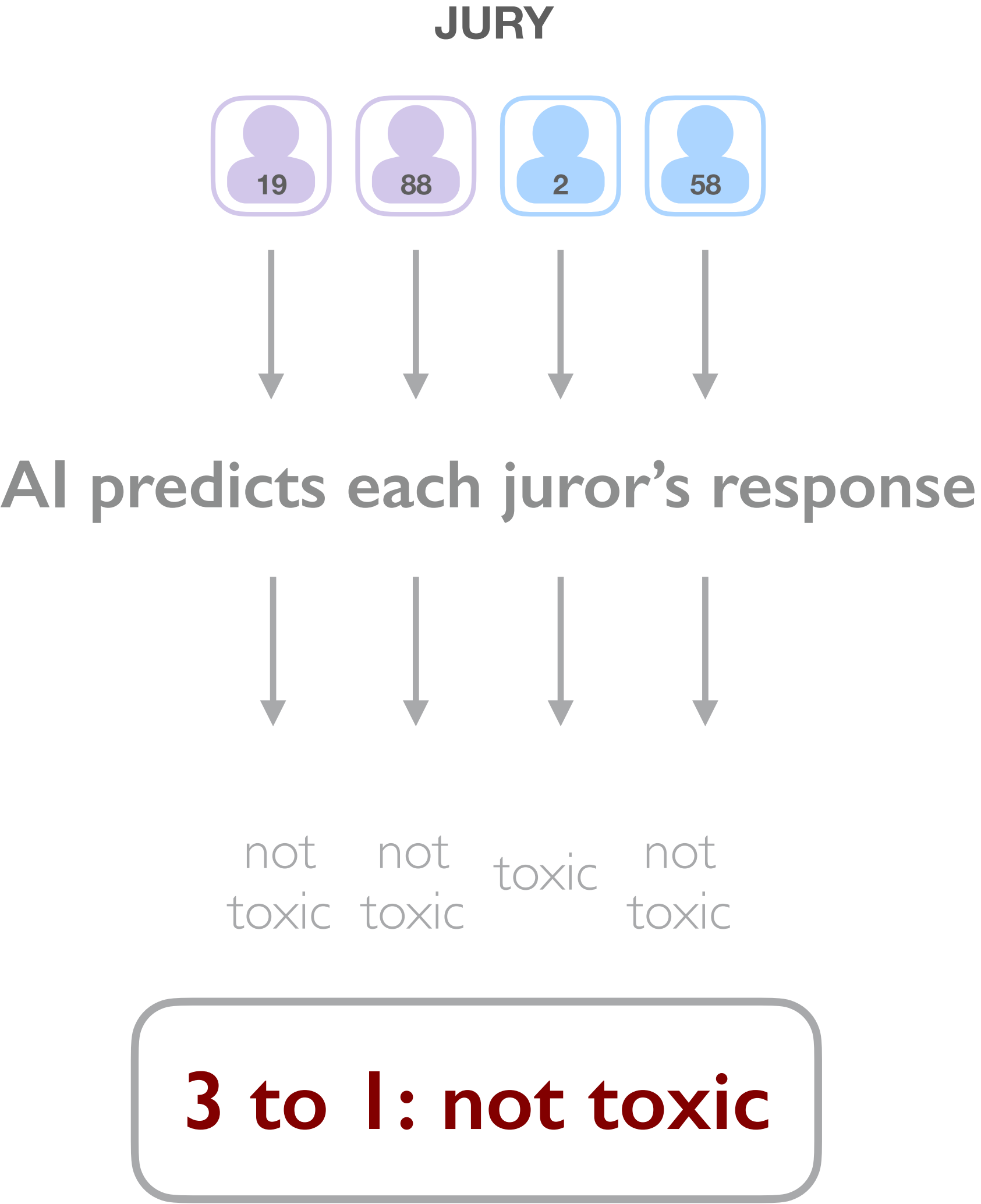
System selects matching annotators from the dataset as jurors



AI predicts how each juror would vote



- A: Liberal
- B: Parents + HS Diploma



What if we change the objective?

[Jia, Lam, et al. 2024]

HCI can articulate **alternative goals and objectives** for these algorithms — what else might we consider?

Let's start with what the social sciences have articulated ➡
[Voelkel et al. 2024]

Science

< BACK TO VOL. 386, NO. 6719

🔒 | RESEARCH ARTICLE | INTERVENTIONS



Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity

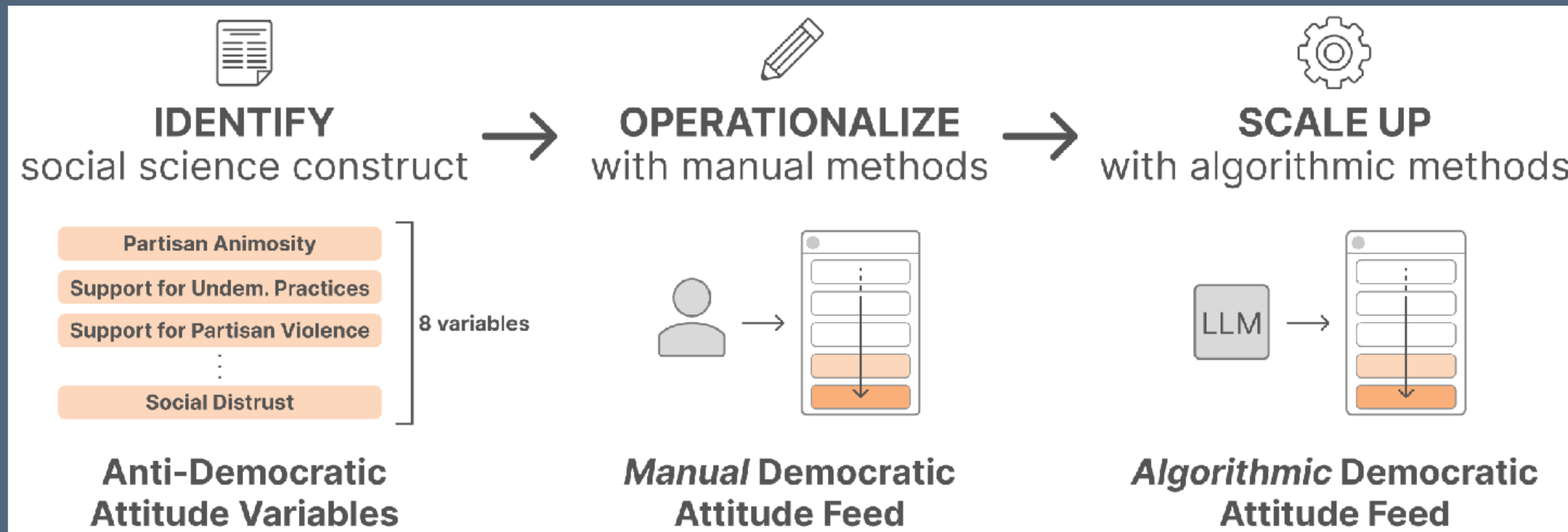
JAN G. VOELKEL , MICHAEL N. STAGNARO, [...], AND ROBB WILLER 

+82 authors

[Authors Info & Affiliations](#)

What if we change the objective?

[Jia, Lam, et al. 2024]



Result: **decreased partisan animosity** in the lab. Would it hold up in the field?

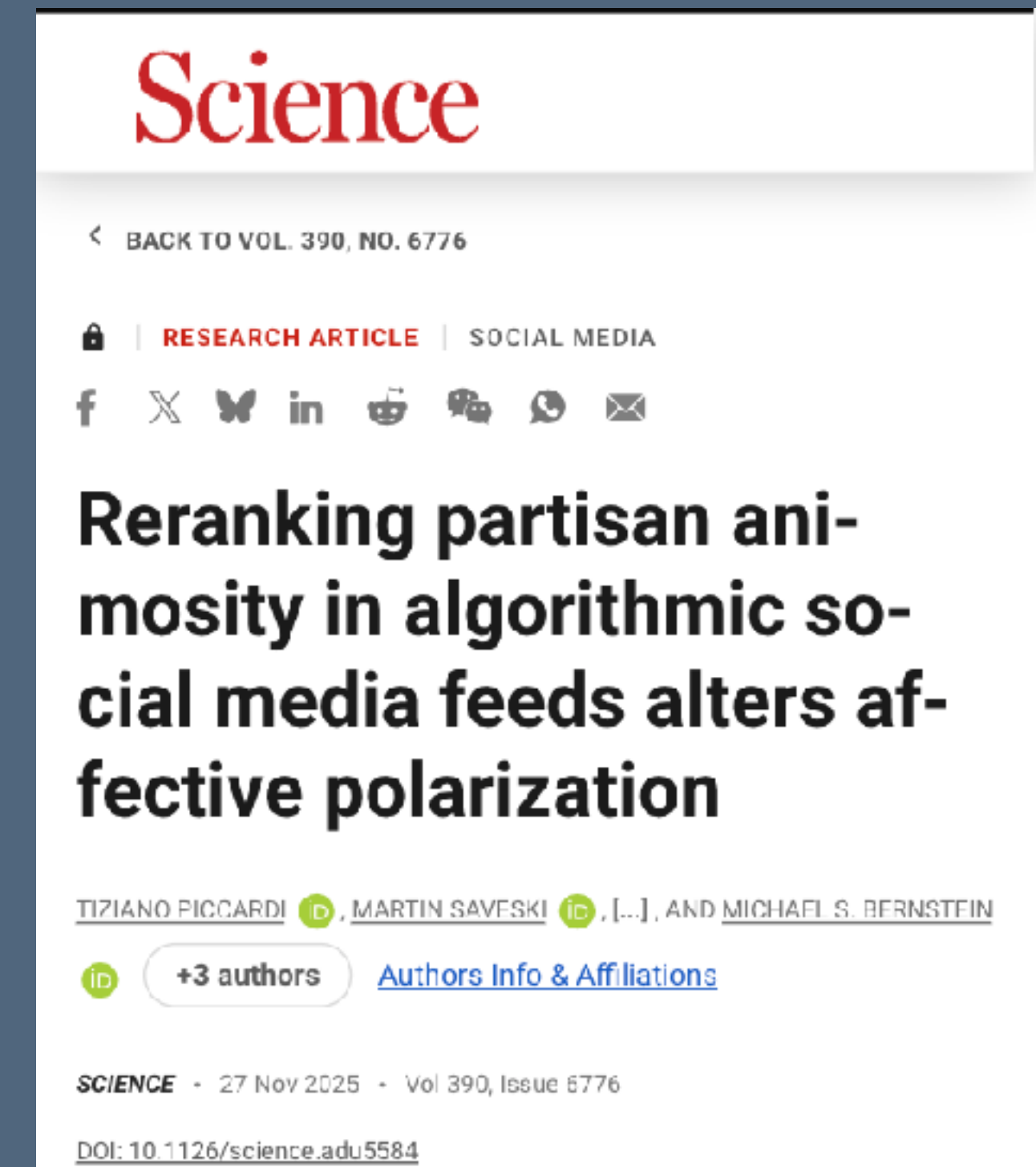
In the field [Piccardi et al. 2025]

A browser extension reranks participants' real X feeds in real time without platform cooperation

1,256 participants during the 2024 US presidential campaign

Reranking changed exposure to posts expressing antidemocratic attitudes and partisan animosity

Result: out-party animosity **shifted by more than 2 points** on a 100-point scale, in both directions, with no detectable difference by party

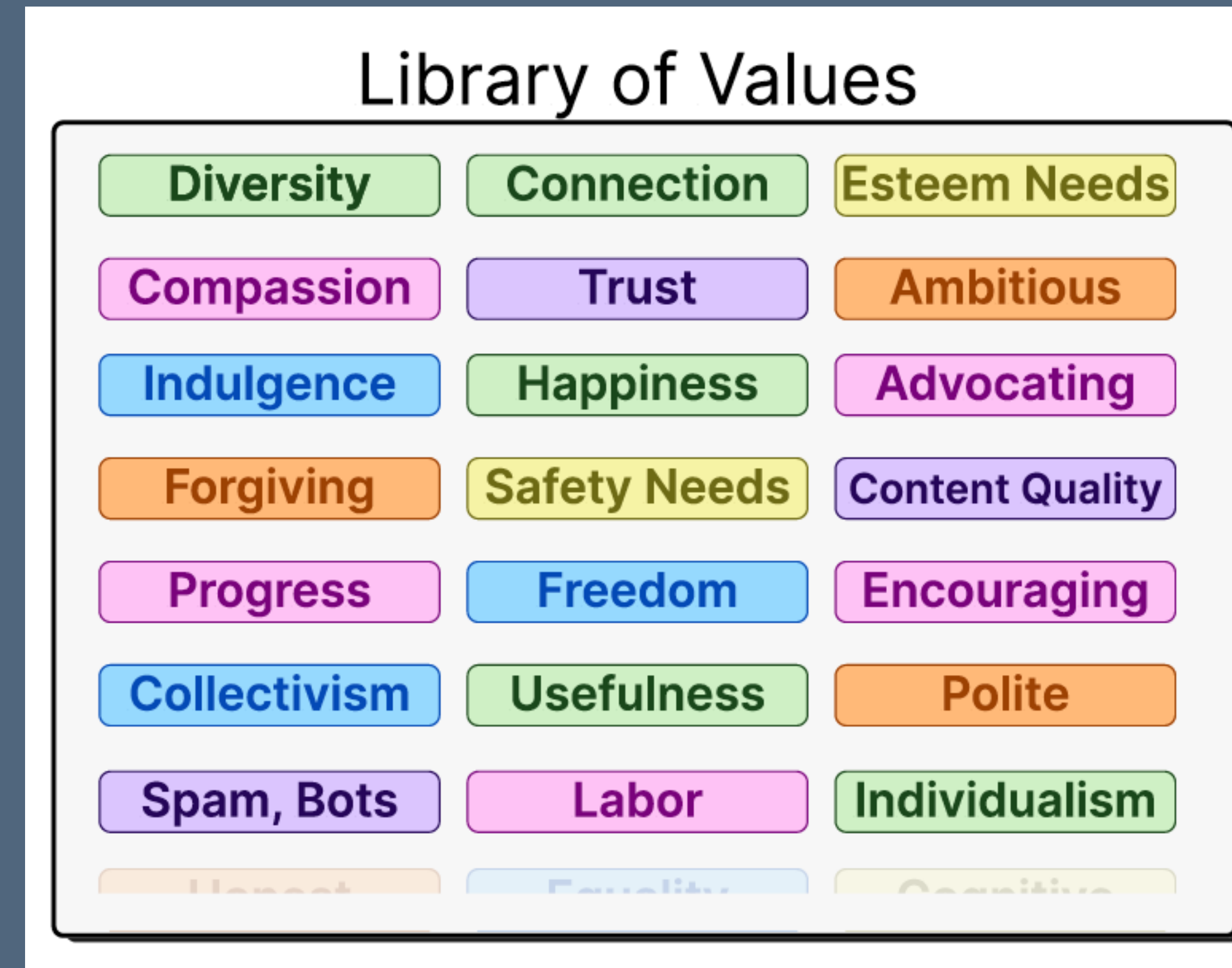


What objectives should we be aiming for?

What do you think we ought to be designing into our algorithms?

What kinds of controls ought end users to have over these algorithms?

What should (and shouldn't) regulation have to say about this kind of alignment?



Summary

Artifacts have politics: the systems we create influence groups and societies, often with undesirable outcomes

Example: gig economy — potential of upward mobility and community social capital, but not currently implemented in a way that unlocks those possibilities

Design approaches focused on marginalized groups, such as **feminist HCI**, center these communities' needs in the design process

Algorithmic systems, not just designed systems, similarly have impact. People struggle to reason about them, and industry struggles to avoid mistakes. But, **modeling human-centered objectives** can help.

References

Agarwal, Dhruv, Mor Naaman, and Aditya Vashistha. "AI suggestions homogenize writing toward Western styles and diminish cultural nuances." Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025.

Angwin, Julia, et al. "Machine bias." ProPublica. 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Bardzell, Shaowen. "Feminist HCI: taking stock and outlining an agenda for design." Proceedings of the SIGCHI conference on human factors in computing systems. 2010.

Barrett, Paul, Justin Hendrix, and Grant Sims. "How tech platforms fuel U.S. political polarization and what government can do about it." Brookings, September 27, 2021.

Bastani, Hamsa, et al. "Generative AI without guardrails can harm learning: Evidence from high school mathematics." Proceedings of the National Academy of Sciences 122.26 (2025): e2422633122.

Bran, Mirel. "Romania's presidential election canceled over TikTok disinformation investigation." Le Monde, December 7, 2024.

Calacci, Dan, and Alex Pentland. "Bargaining with the black-box: Designing and deploying worker-centric tools to audit algorithmic management." Proceedings of the ACM on Human-Computer Interaction 6.CSCW2 (2022): 1-24.

Calacci, Dan, et al. "FairFare: A tool for crowdsourcing rideshare data to empower labor organizers." ACM Transactions on Computer-Human Interaction 33.1 (2025).

Chan, Kelvin. "Musk's AI chatbot faces global backlash over sexualized images of women and children." Associated Press, January 13, 2026.

Costanza-Chock, Sasha. Design Justice: Community-Led Practices to Build the Worlds We Need. MIT Press, 2020.

Costello, Thomas H., Gordon Pennycook, and David G. Rand. "Durably reducing conspiracy beliefs through dialogues with AI." Science 385.6714 (2024): eadq1814.

References

Dillahunt, Tawanna R., and Amelia R. Malone. "The promise of the sharing economy among disadvantaged communities." Proceedings of the 33rd annual ACM conference on human factors in computing systems. 2015.

Dombrowski, Lynn, Ellie Harmon, and Sarah Fox. "Social justice-oriented interaction design: Outlining key design strategies and commitments." Proceedings of the 2016 ACM Conference on Designing Interactive Systems. 2016.

Dubal, Veena. "On algorithmic wage discrimination." Columbia Law Review 123.7 (2023): 1929-1992.

European Union. "Directive (EU) 2024/2831 on improving working conditions in platform work." Official Journal of the European Union, November 11, 2024.

Gordon, Mitchell L., et al. "Jury learning: Integrating dissenting voices into machine learning models." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022.

Gray, Mary L., and Siddharth Suri. Ghost work: How to stop Silicon Valley from building a new global underclass. Eamon Dolan Books, 2019.

Groves, Lara, et al. "Auditing work: Exploring the New York City algorithmic bias audit regime." Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. 2024.

Hackenburg, Kobi, et al. "The levers of political persuasion with conversational artificial intelligence." Science 390.6777 (2025): eaea3884.

Hackenburg, Kobi, Luke Hewitt, Caroline Wagner, Christopher Summerfield, and Ben M. Tappin. "AI can persuade people to take political actions." Proceedings of the National Academy of Sciences 123.42 (2026): e2611368123.

Hara, Kotaro, et al. "A data-driven analysis of workers' earnings on Amazon Mechanical Turk." Proceedings of the 2018 CHI conference on human factors in computing systems. 2018.

References

Harrington, Christina, et al. "'It's kind of like code-switching': Black older adults' experiences with a voice assistant for health information seeking." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022.

Hayes, Gillian R. "The relationship of action research to human-computer interaction." ACM Transactions on Computer-Human Interaction (TOCHI) 18.3 (2011): 1-20.

Hill, Kashmir, Kalley Huang, and Mike Isaac. "Meta plans to add facial recognition technology to its smart glasses." The New York Times, February 13, 2026.

Hofmann, Valentin, et al. "AI generates covertly racist decisions about people based on their dialect." Nature 633.8028 (2024): 147-154.

Holstein, Kenneth, et al. "Improving fairness in machine learning systems: What do industry practitioners need?." Proceedings of the 2019 CHI conference on human factors in computing systems. 2019.

Horwitz, Jeff. "Meta's flirty AI chatbot invited a retiree to New York. He never made it home." Reuters, August 14, 2025.

Ibrahim, Lujain, et al. "Sycophantic AI makes human interaction feel more effortful and less satisfying over time." arXiv preprint arXiv:2605.07912 (2026).

Irani, Lilly C., and M. Six Silberman. "Turkopticon: Interrupting worker invisibility in amazon mechanical turk." Proceedings of the SIGCHI conference on human factors in computing systems. 2013.

Jackson, Steven J., Tarleton Gillespie, and Sandy Payette. "The policy knot: Re-integrating policy, practice and design in CSCW studies of social computing." Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. 2014.

Jakesch, Maurice, et al. "Co-writing with opinionated language models affects users' views." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.

References

Jia, Chenyan, Michelle S. Lam, Minh Chau Mai, Jeffrey T. Hancock, and Michael S. Bernstein. "Embedding democratic values into social media AIs via societal objective functions." Proceedings of the ACM on Human-Computer Interaction 8.CSCWI (2024): 1-36.

Kittur, Aniket, et al. "The future of crowd work." Proceedings of the 2013 conference on Computer supported cooperative work. 2013.

Knutsson, Kurt. "Tom Brady, Paris Hilton, Snoop Dogg and Kendall Jenner change their names for AI." Fox News, October 6, 2023.

Kolluri, Akaash, et al. "Alexandria: A library of pluralistic values for realtime re-ranking of social media feeds." Proceedings of the International AAAI Conference on Web and Social Media (ICWSM). 2026.

Lee, Min Kyung, et al. "Working with machines: The impact of algorithmic and data-driven management on human workers." Proceedings of the 33rd annual ACM conference on human factors in computing systems. 2015.

Lin, Hause, et al. "Persuading voters using human–artificial intelligence dialogues." Nature 648 (2025): 394-401.

Linxen, Sebastian, et al. "How WEIRD is CHI?" Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021.

Louie, Ryan, et al. "Can LLM-simulated practice and feedback upskill human counselors? A randomized study with 90+ novice counselors." Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. 2026.

Martin, David, et al. "Being a turker." Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. 2014.

Metaxa, Danaë, et al. "Auditing algorithms: Understanding algorithmic systems from the outside in." Foundations and Trends® in Human–Computer Interaction 14.4 (2021): 272-344.

References

Nature. "Tweaking Facebook feeds is no easy fix for polarization, studies find." Nature News, July 27, 2023.

Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." Science 366.6464 (2019): 447-453.

Piccardi, Tiziano, et al. "Reranking partisan animosity in algorithmic social media feeds alters affective polarization." Science 390.6776 (2025): eadu5584.

Quigley, Liam. "Regulators say DoorDash and Uber Eats chiseled NYC delivery workers out of \$550M in tips." Gothamist, January 13, 2026.

Reich, Rob, Mehran Sahami, and Jeremy M. Weinstein. System Error: Where Big Tech Went Wrong and How We Can Reboot. Harper, 2021.

Roose, Kevin. "Don't ban ChatGPT in schools. Teach with it." The New York Times, January 12, 2023.

Rose-Stockwell, Tobias. Outrage Machine: How Tech Amplifies Discontent, Disrupts Democracy—And What We Can Do About It. Legacy Lit, 2023.

Salehi, Niloufar, et al. "We are dynamo: Overcoming stalling and friction in collective action for crowd workers." Proceedings of the 33rd annual ACM conference on human factors in computing systems. 2015.

Sandvig, Christian, et al. "Auditing algorithms: Research methods for detecting discrimination on internet platforms." Data and discrimination: converting critical concerns into productive inquiry 22.2014 (2014): 4349-4357.

Sweeney, Latanya. "Discrimination in online ad delivery." Communications of the ACM 56.5 (2013): 44-54.

Törnberg, Petter. "How digital media drive affective polarization through partisan sorting." Proceedings of the National Academy of Sciences 119.42 (2022): e2207159119.

References

Umbach, Rebecca, et al. "Non-consensual synthetic intimate imagery: Prevalence, attitudes, and knowledge in 10 countries." Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024.

Voelkel, Jan G., et al. "Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity." Science 386.6719 (2024): eadh4764.

Walsh, James D. "Everyone is cheating their way through college." New York Magazine, May 2025.

Wang, Zijie J., et al. "Farsight: Fostering responsible AI awareness during AI application prototyping." Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024.

Williams-Ceci, Sterling, et al. "Biased AI writing assistants shift users' attitudes on societal issues." Science Advances 12.11 (2026): eadw5578.

Winner, Langdon. "Do artifacts have politics?" Daedalus 109.1 (1980): 121-136.

Wright, Lucas, et al. "Null compliance: NYC Local Law 144 and the challenges of algorithm accountability." Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency. 2024.

Yang, Qian, et al. "The future of HCI-policy collaboration." Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024.

Zhang, Angie, et al. "Algorithmic management reimaged for workers and by workers: Centering worker well-being in gig work." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022.

Zhang, Renwen, et al. "The dark side of AI companionship: A taxonomy of harmful algorithmic behaviors in human-AI relationships." Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025.

Zhang, Yutong, Dora Zhao, Jeffrey T. Hancock, Robert E. Kraut, and Diyi Yang. "Interaction with AI companions and psychological well-being." Nature Human Behaviour (2026).

Zhang, Yutong, et al. "Living with AI companions: Sustained AI companionship predicts lower well-being through lower human interaction." arXiv preprint arXiv:2609.07243 (2026).

Zhu, Haiyi, Bowen Yu, Aaron Halfaker, and Loren Terveen. "Value-sensitive algorithm design: Method, case study, and lessons." Proceedings of the ACM on Human-Computer Interaction 2.CSCW (2018): 1-23.