



AI  
ACCOUNTABILITY  
LAB



AI for Good

# AI FOR SOCIAL GOOD: THE NEW FACE OF TECHNOSOLUTIONISM

Abeba Birhane

Founder & PI AIAL

School of Computer Science and Statistics

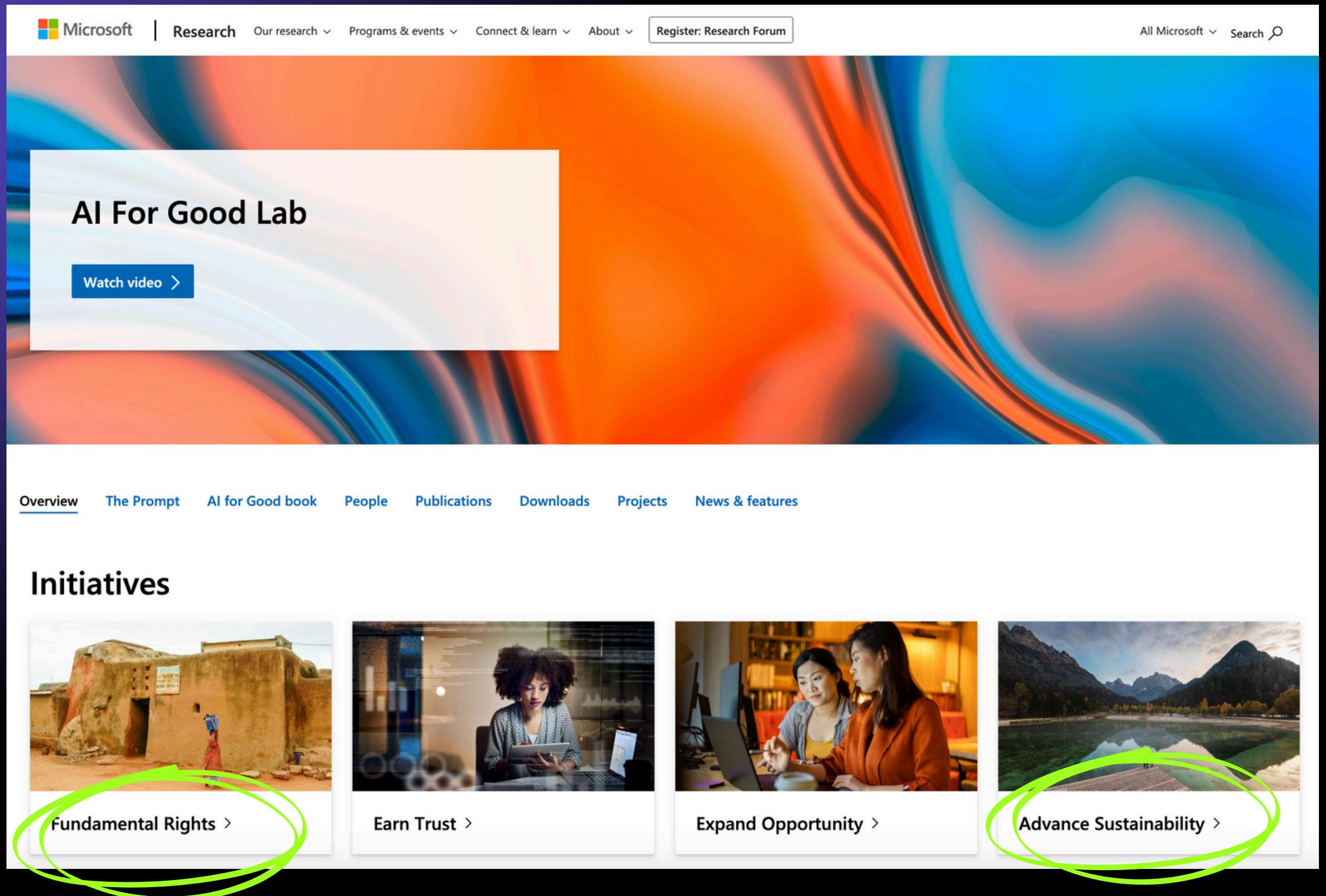
Trinity College Dublin

Jul. 8, 2025



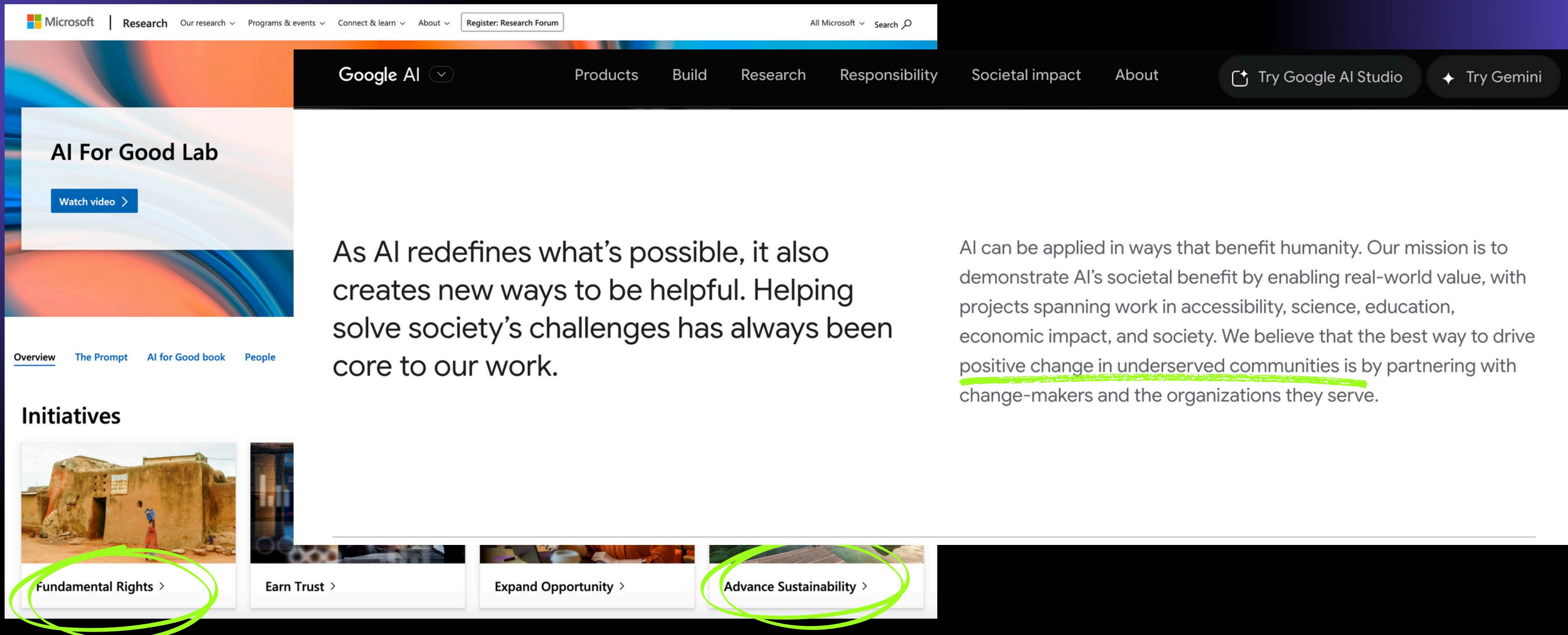
@ABEBA.BSKY.SOCIAL

# AI FOR SOCIAL GOOD: A TREND



The screenshot shows the Microsoft AI For Good Lab website. The header includes the Microsoft logo, a navigation bar with 'Research', 'Our research', 'Programs & events', 'Connect & learn', 'About', 'Register: Research Forum', 'All Microsoft', and a search bar. The main section features a large, colorful abstract background image. On the left, a white box contains the text 'AI For Good Lab' and a 'Watch video >' button. Below this, a navigation menu lists 'Overview', 'The Prompt', 'AI for Good book', 'People', 'Publications', 'Downloads', 'Projects', and 'News & features'. The 'Overview' item is underlined. The 'Initiatives' section displays five cards with images: 1. A woman in a red dress standing outside a mud-brick building, with a green circle highlighting the 'Fundamental Rights >' text below it. 2. Two women working on laptops in an office, with a green circle highlighting the 'Earn Trust >' text below it. 3. Two women working on laptops in an office, with a green circle highlighting the 'Expand Opportunity >' text below it. 4. A landscape of mountains and a lake, with a green circle highlighting the 'Advance Sustainability >' text below it.

# AI FOR SOCIAL GOOD: A TREND



Microsoft | Research Our research ▾ Programs & events ▾ Connect & learn ▾ About ▾ Register: Research Forum All Microsoft ▾ Search 

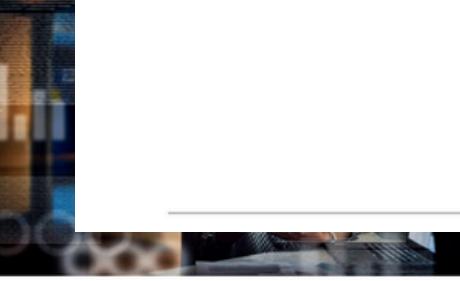
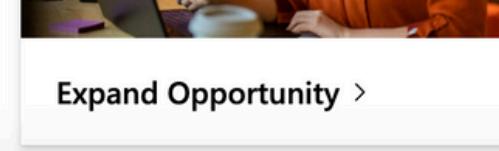
Google AI ▾ Products Build Research Responsibility Societal impact About  Try Google AI Studio  Try Gemini

## AI For Good Lab

Watch video >

Overview The Prompt AI for Good book People

### Initiatives

 Fundamental Rights >  Earn Trust >  Expand Opportunity >  Advance Sustainability >

As AI redefines what's possible, it also creates new ways to be helpful. Helping solve society's challenges has always been core to our work.

AI can be applied in ways that benefit humanity. Our mission is to demonstrate AI's societal benefit by enabling real-world value, with projects spanning work in accessibility, science, education, economic impact, and society. We believe that the best way to drive positive change in underserved communities is by partnering with change-makers and the organizations they serve.

## Working Paper

# Does the United Nations need agents? Testing the role of AI agent generated personas in humanitarian action

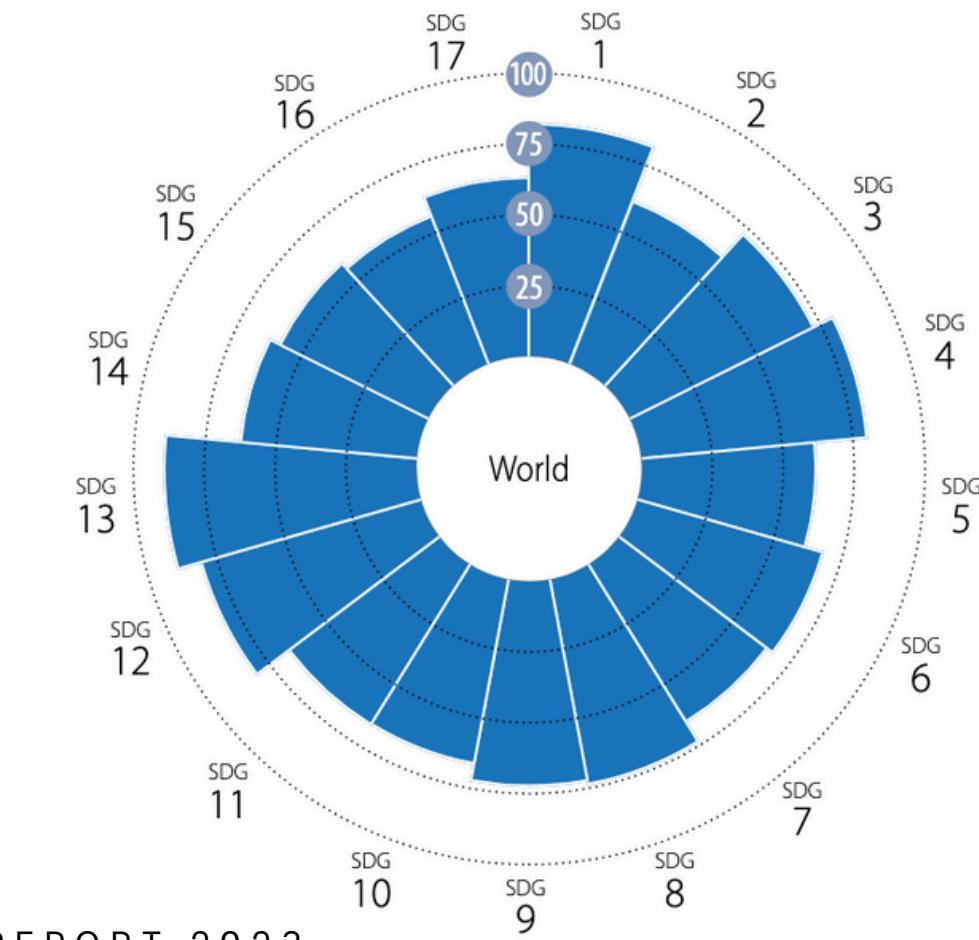
Amina simulates a refugee living in Chad's Metche camp, while Abdalla represents a Rapid Support Forces combatant in eastern Sudan. Both are designed to authentically represent their respective populations through specifically curated knowledge bases - which in turn provide the foundation for the synthetically generated opinions. **Both personas are available for public interaction and conversation at [askamina.ai](http://askamina.ai).**

UN 2025

## Overall Performance



## Average Performance by SDG



## SDG Dashboard and Trends



- Major challenges
- Significant challenges
- Challenges remain
- SDG achieved
- Information unavailable
- ⬇ Decreasing
- ➡ Stagnating
- ↗ Moderately improving
- ↑ On track or maintaining SDG achievement
- Information unavailable



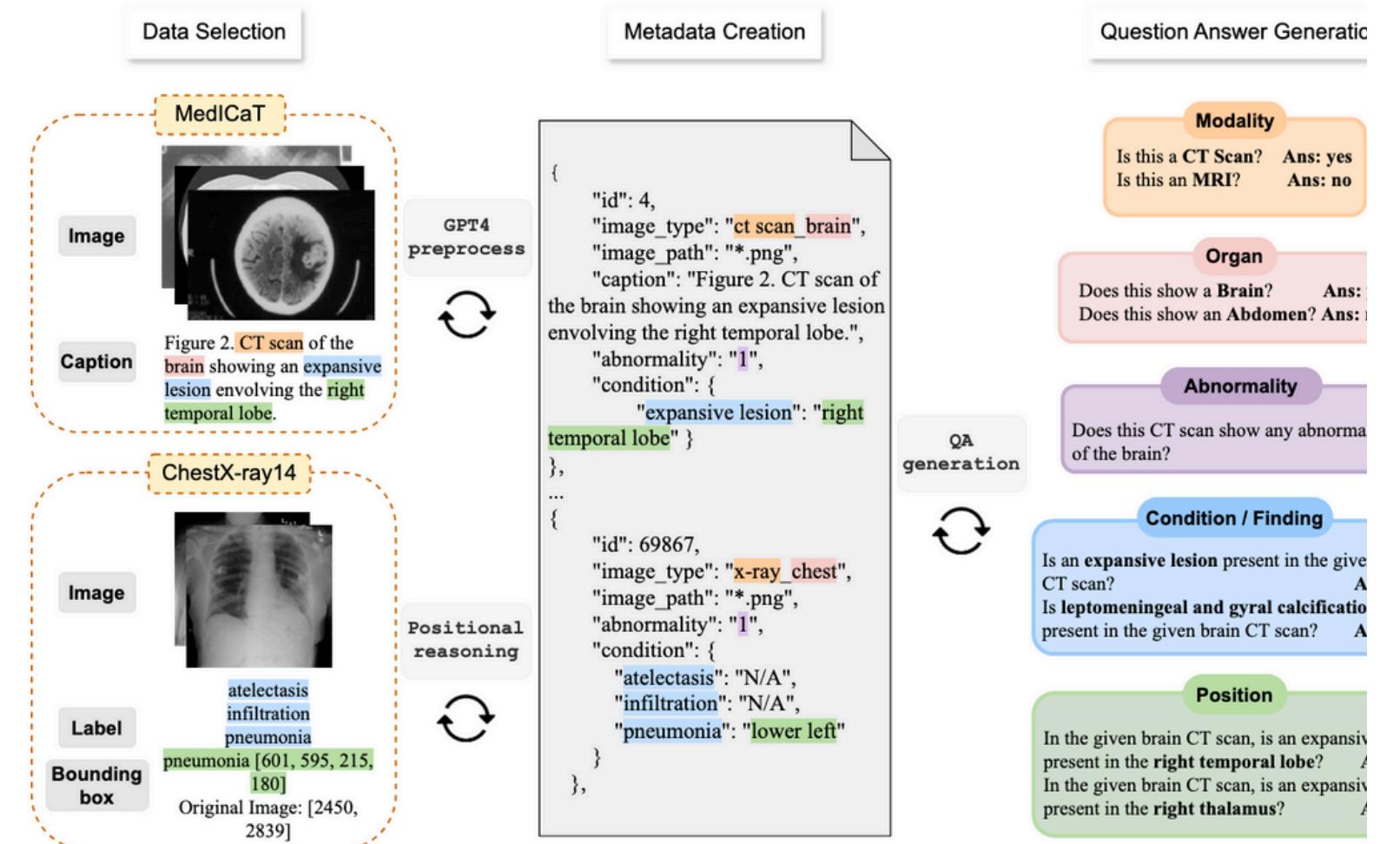


Figure 3: Flow diagram of the ProbMed data curation process. Two comprehensive biomedical datasets were utilized to collect source data and construct a metadata file, enabling the automatic generation of high-quality question-answer pairs for the ProbMed dataset.

MAGESH ET AL., 2024

## Worse than Random? An Embarrassingly Simple Probing Evaluation of Large Multimodal Models in Medical VQA

**Qianqi Yan**  
University of California, Santa Cruz  
qyan79@ucsc.edu

**Xuehai He**  
University of California, Santa Cruz  
xhe89@ucsc.edu

**Xiang Yue**  
Carnegie Mellon University  
xyue2@andrew.cmu.edu

**Xin Eric Wang**  
University of California, Santa Cruz  
xwang366@ucsc.edu

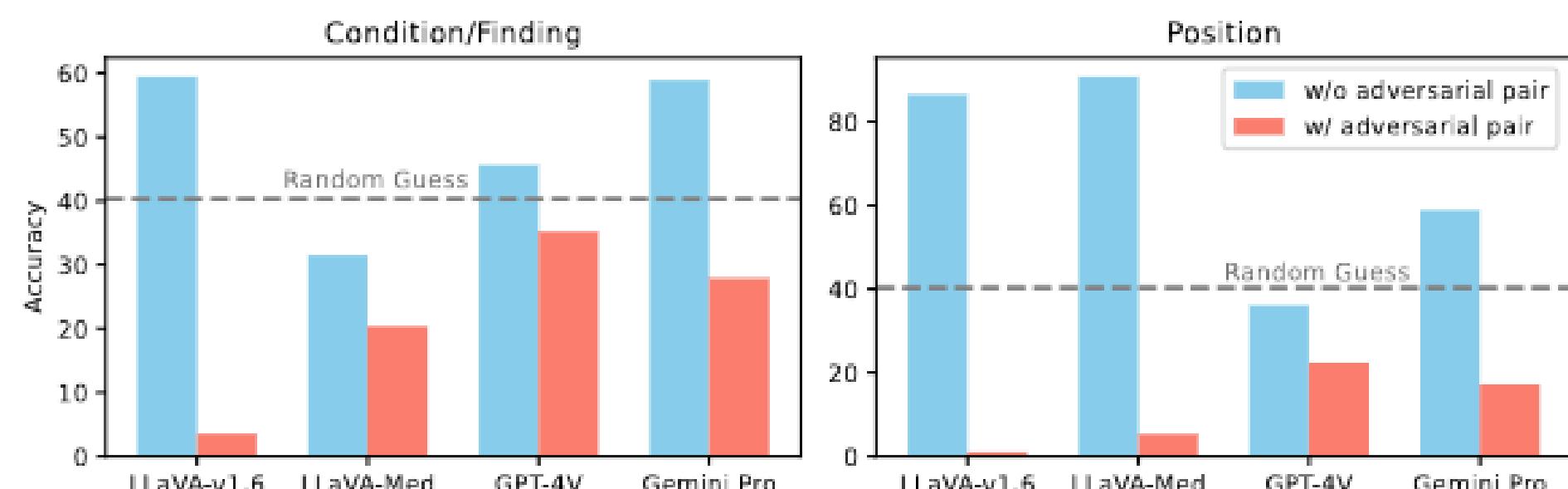


Figure 1: Accuracy of four LMMs on two types of specialized questions in medical diagnoses, with and without adversarial pairs. The significant drop in accuracy with adversarial pairs highlights the models' unreliability in handling medical diagnoses.

# AI TECHNOLOGIES OFTEN FAIL TO LIVE UP TO THEIR CLAIMS

Anonymised Model Name	Model A	Model B
Model name	Mistral-7B	Mistral-7B
Licence	Apache 2.0	Apache 2.0
Commercial use	Yes	Yes
Base model	Mistral-7B	Mistral-7B
MMLU score <sup>3</sup>	64.16	55.5
Average <sup>4</sup>	60.97	55.5
Context length <sup>5</sup>	8k	8k
Model size <sup>6</sup>	7B	7B

Criteria/Rating	AI generated summary scores - aggregated all 5 submissions	Human summary scores - aggregated all 5 submissions
Coherency/Consistency	10	12
References to ASIC	5	15
Identifies recommendations on how conflicts of interest should be regulated	5	8
References to more regulation of auditors/consultants	6	11
Length	9	15
<b>Total (out of 75)</b>	<b>35</b>	<b>61</b>
<b>Percentage of total</b>	<b>47%</b>	<b>81%</b>

Table 4 - ASIC Scoring Summary by Assessment Criteria. The maximum score per summary was 15.

The assessment rubric scores for each submission were:

Submission	AI generated summary scores (Total)	Human summary scores (Total)
Institute of Public Accountants	6	15
KPMG Australia	8	9
ATO	8	15
Dr Kelli Larson	5	10
The Australia Institute	8	12
<b>Total of documents (out of 75)</b>	<b>35</b>	<b>61</b>
<b>Percentage of total</b>	<b>47%</b>	<b>81%</b>

Table 5 - ASIC Scoring Summary by Submission. The maximum score per summary was 15.

Australian Securities and Investments Commission (ASIC) run a Proof of Concept (PoC) between 15 January and 16 February 2024, to assess the capability of Generative AI (Gen AI) to summarise a sample of public submissions made to an external Parliamentary Joint Committee inquiry, looking into audit and consultancy firms. (Wilson, 2024)

# AI TECHNOLOGIES OFTEN FAIL TO LIVE UP TO THEIR CLAIMS

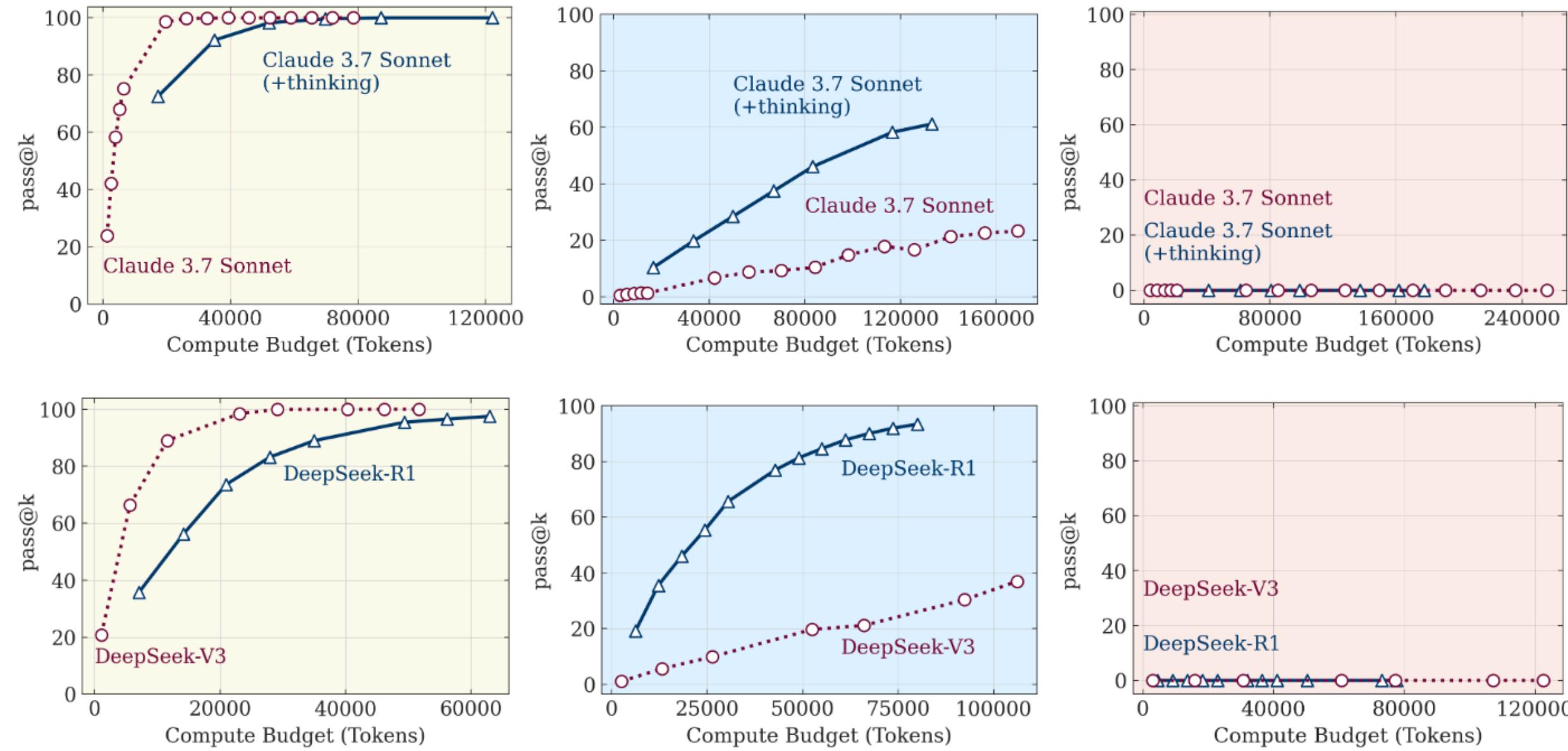
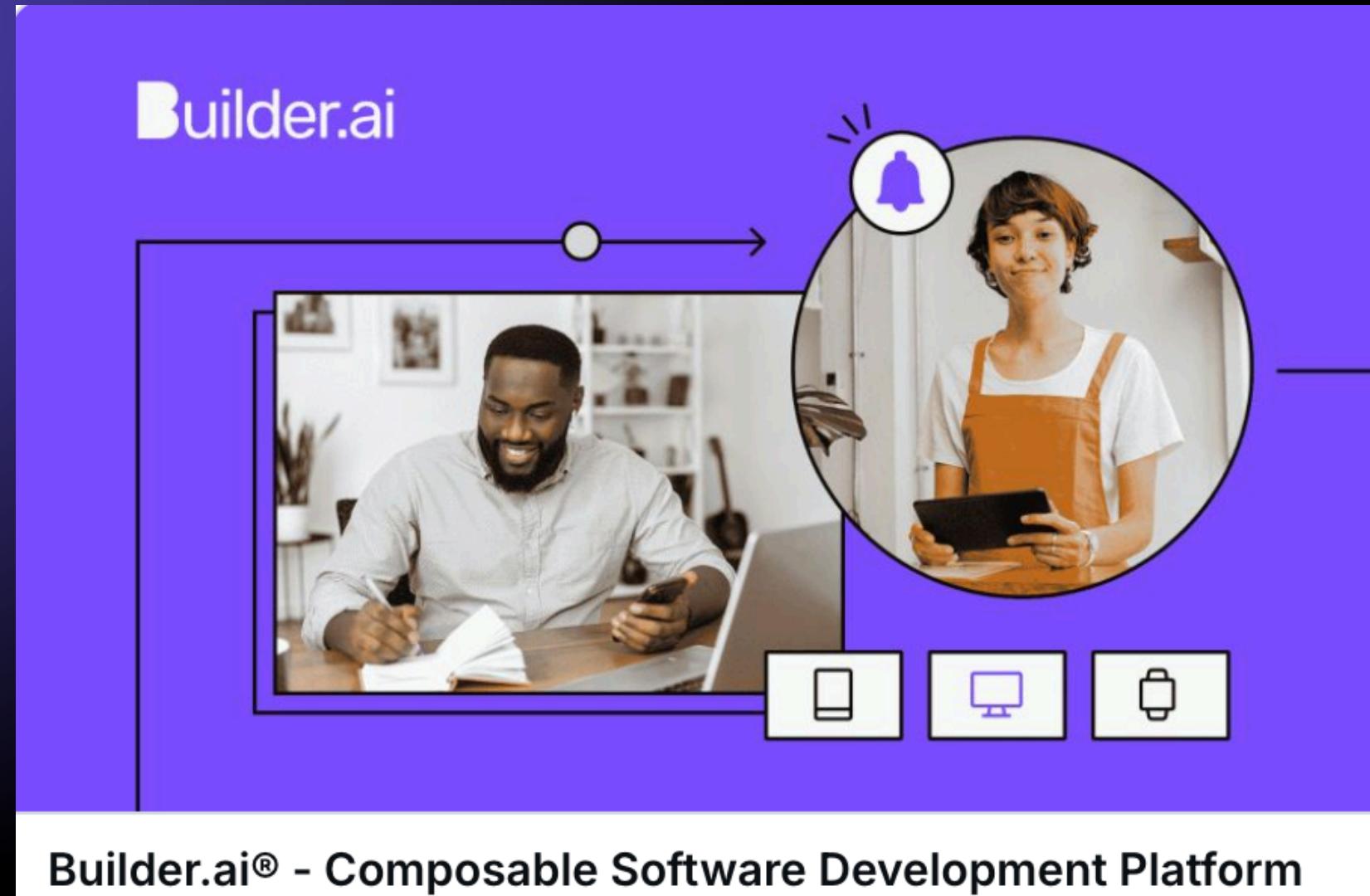


Figure 5: Pass@k performance of thinking vs. non-thinking models across equivalent compute budgets in puzzle environments of **low**, **medium**, and **high** complexity. Non-thinking models excel in simple problems, thinking models show advantages at medium complexity, while both approaches fail at high complexity regardless of compute allocation.



## AI TECHNOLOGIES OFTEN FAIL TO LIVE UP TO THEIR CLAIMS



Builder.ai has collapsed under the weight of several deceptions, ending the rise of a British startup once heralded as a pioneer in democratizing software development. Valued at \$1.5 billion in 2023 with backing from Microsoft and Qatar's sovereign wealth fund, the company filed for bankruptcy protection this week after revelations that its "AI-powered" platform relied on hundreds of engineers in India while executives allegedly orchestrated a years-long financial deception.

For eight years, Builder.ai marketed its "Natasha" AI system as a fully autonomous tool that could build software "as easily as ordering pizza." However, internal documents and employee accounts reviewed by Bloomberg paint a sharply different picture. Engineers in Noida and Bangalore manually coded client projects while being instructed to mimic AI-generated responses.

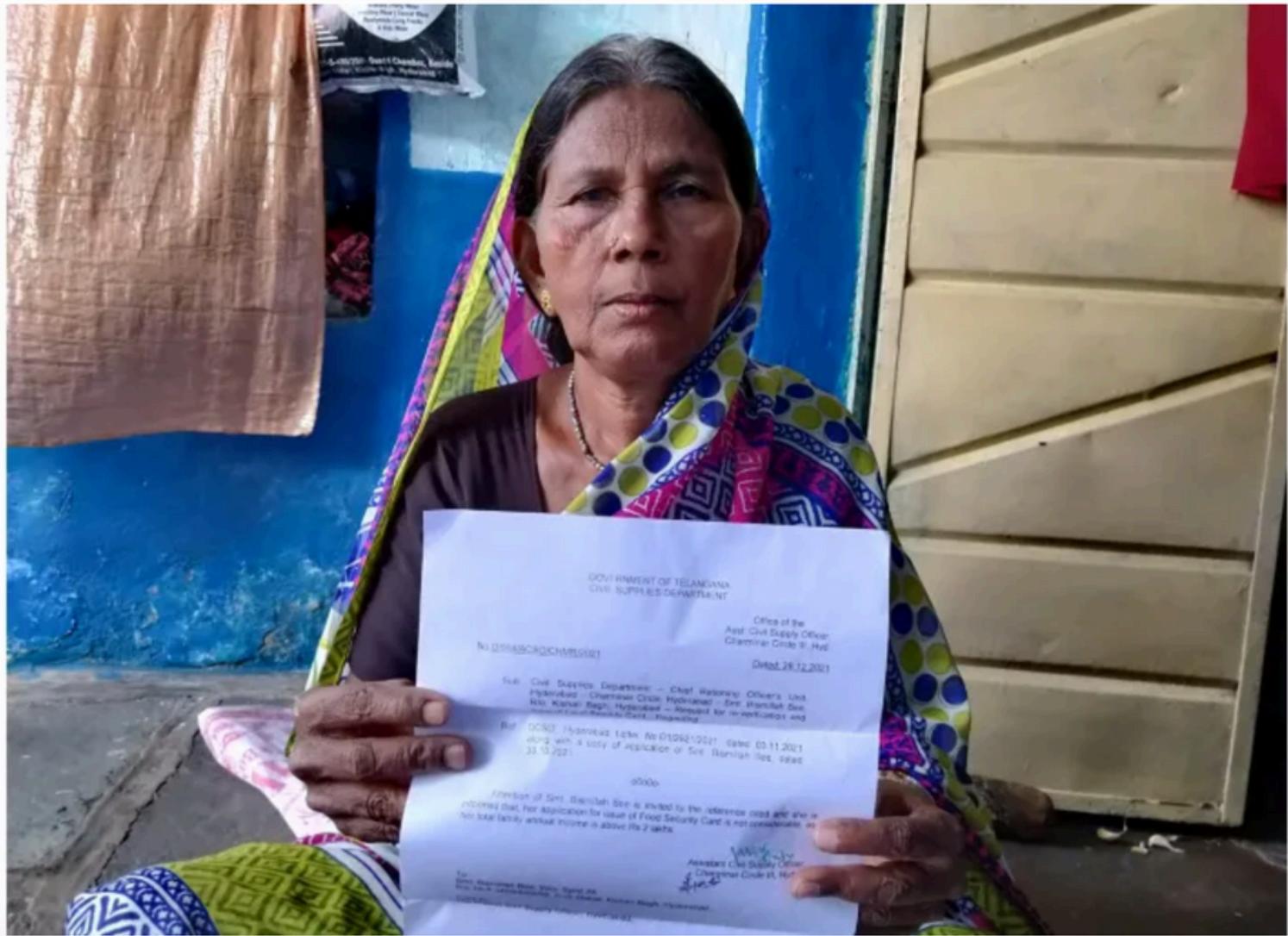
"We were told to never mention our location or use Indian English phrases," said a former Bangalore engineer.

The team timed updates to UK business hours to maintain the illusion of automation.



## How an algorithm denied food to thousands of poor in India's Telangana

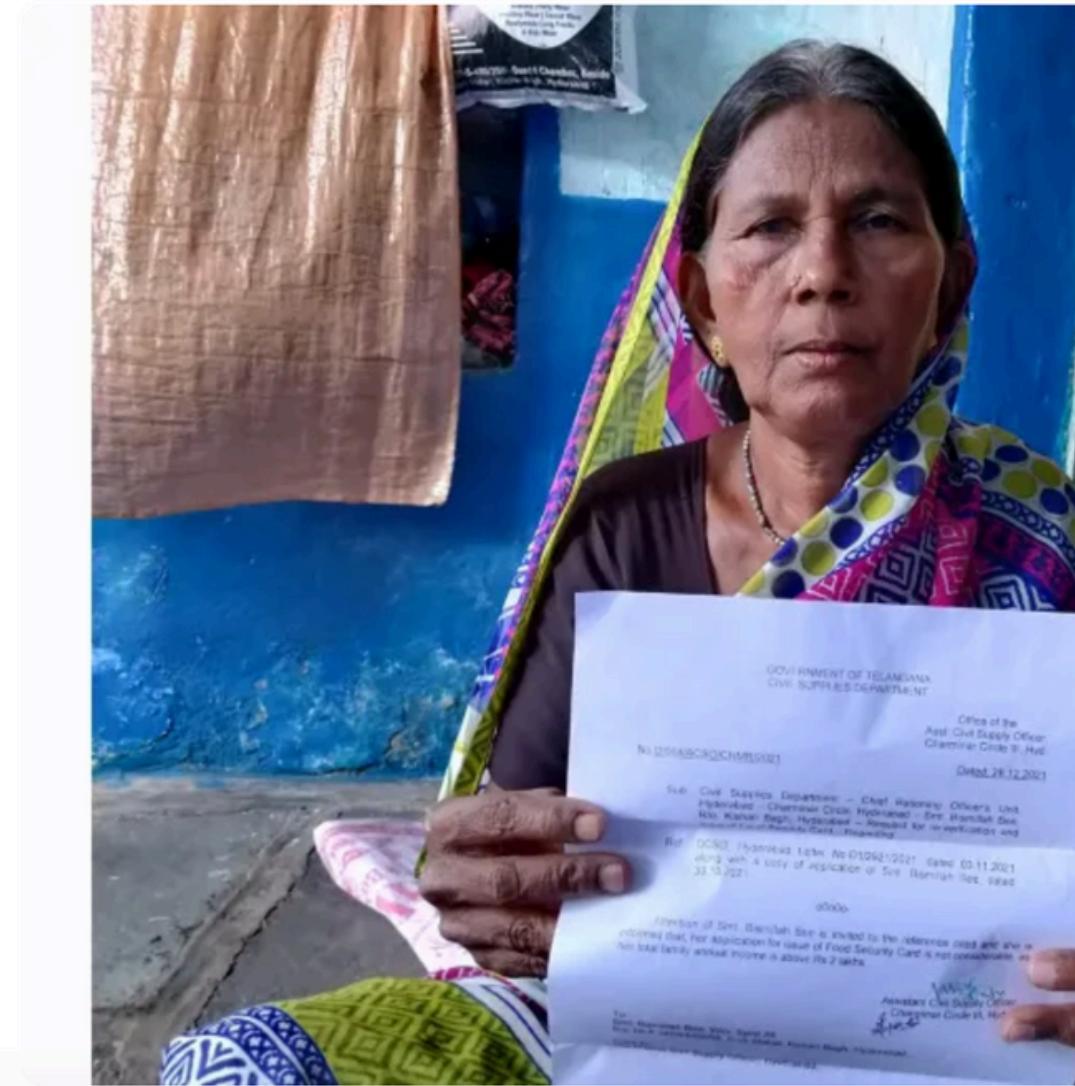
*It adopted AI in welfare schemes to weed out ineligible ones, but has wrongfully removed thousands of legitimate ones.*



ALJAZEERA. JAN 2024

## How an algorithm denied food to thousands of poor in India's

*It adopted AI in welfare schemes to weed out in wrongfully removed thousands of legitimate or*



The Asahi Shimbun | Asia & Japan Watch

Search

HOME What's New National Report Politics Business Asia & World Sci & Tech Culture

地震速報

詳細へ

15時26分頃、鹿児島県鹿児島十島村で最大震度5弱の地震がありました。

The Asahi Shimbun > Sci & Tech > article

### AI designed to spot child abuse risks delayed for inaccuracies

By YUKI KAWANO/ Staff Writer

March 6, 2025 at 17:36 JST

[Share](#) [Tweet](#) [Print](#)

list

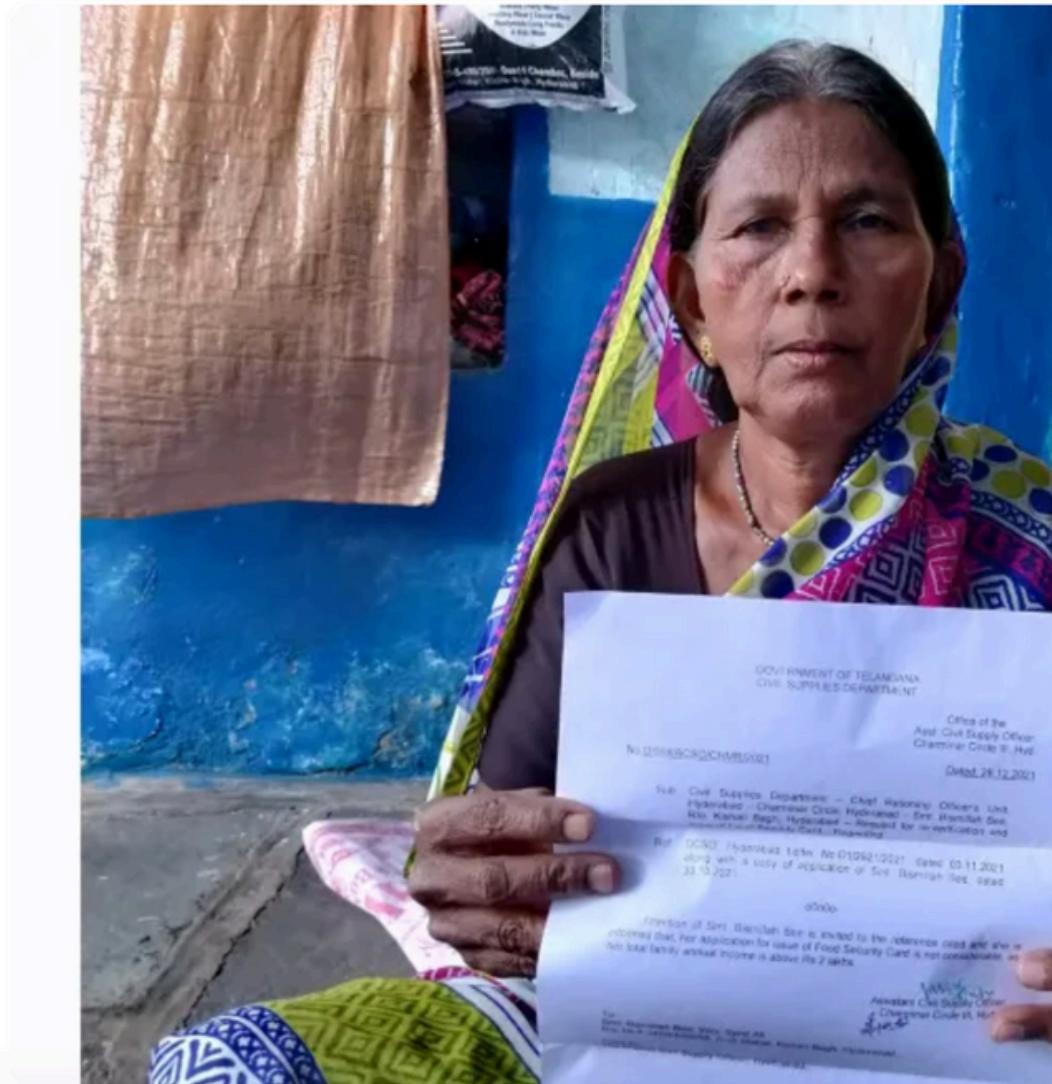


ALJAZEERA. JAN 2024

THE ASAHI SHIMBUN. MAR 2025

## How an algorithm denied food to thousands of poor in India's

*It adopted AI in welfare schemes to weed out inaccuracy. It wrongfully removed thousands of legitimate or*



The Asahi Shimbun | Asia & Japan Watch

HOME What's New National Report Politics Business Asia & World Sci & Tech Culture

地震速報

詳細へ

15時26分頃、鹿児島県鹿児島十島村で最大震度5弱の地震がありました。

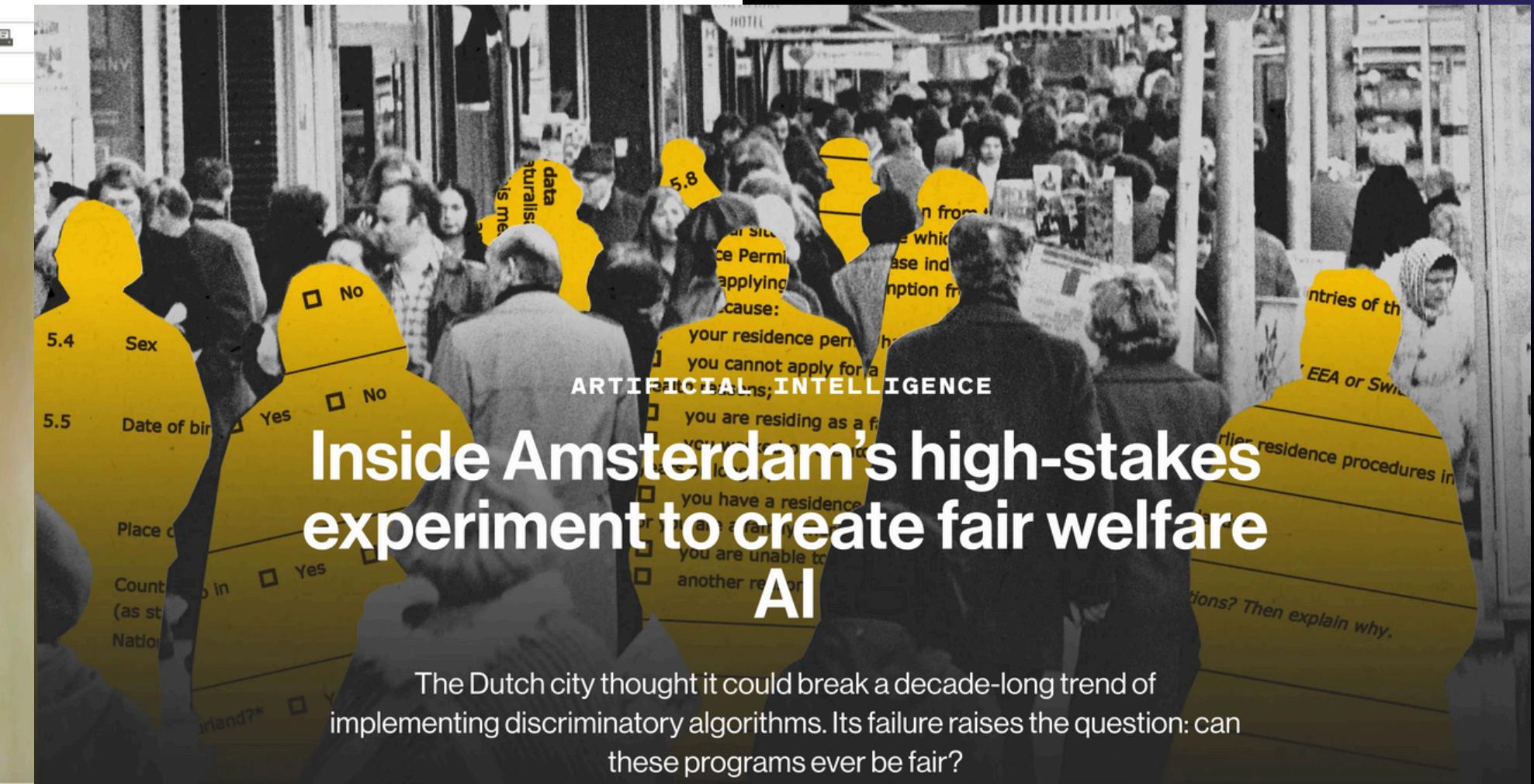
The Asahi Shimbun > Sci & Tech > article

### AI designed to spot child abuse risks delayed for inaccuracies

By YUKI KAWANO/ Staff Writer

March 6, 2025 at 17:36 JST

Share Tweet



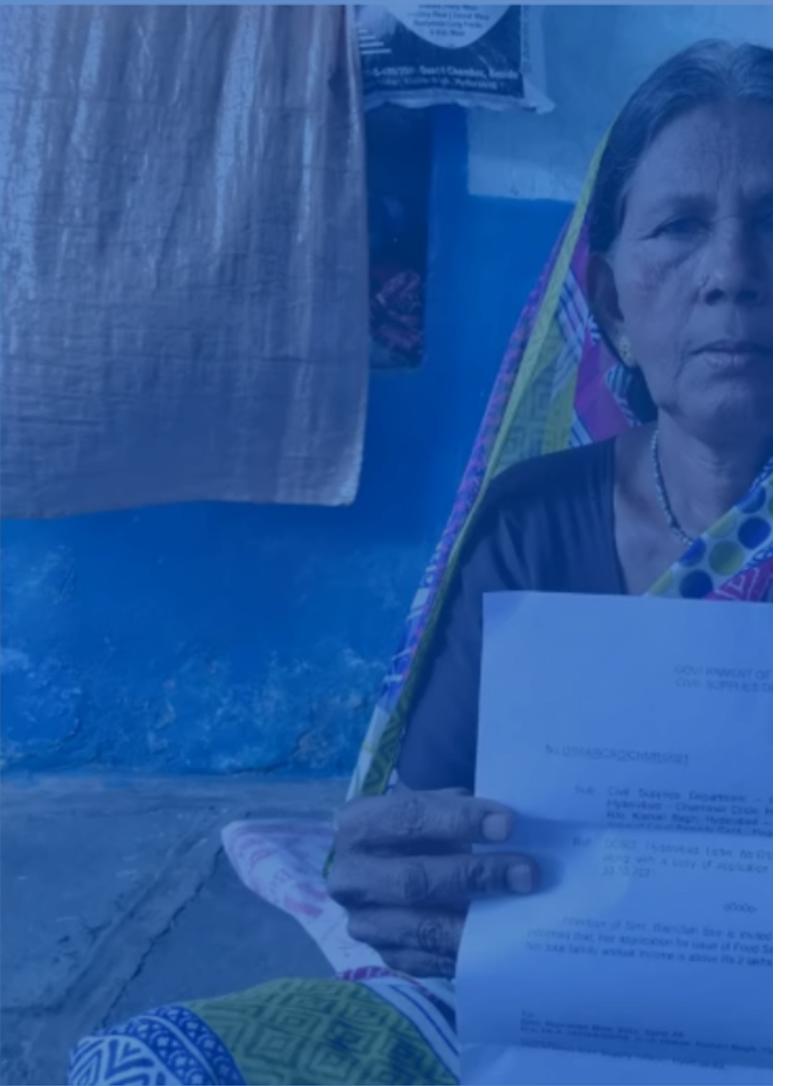
ALJAZEERA. JAN 2024

THE ASAHI SHIMBUN. MAR 2025

MIT TECHNOLOGY REVIEW, LIGHTHOUSE REPORTS, AND TROUW. JUN 2025

## How an algorithm denied thousands of poor in India

*It adopted AI in welfare schemes to update records. It wrongfully removed thousands of legitimate claimants.*



Amsterdam's model used 15 criteria to score welfare applications.

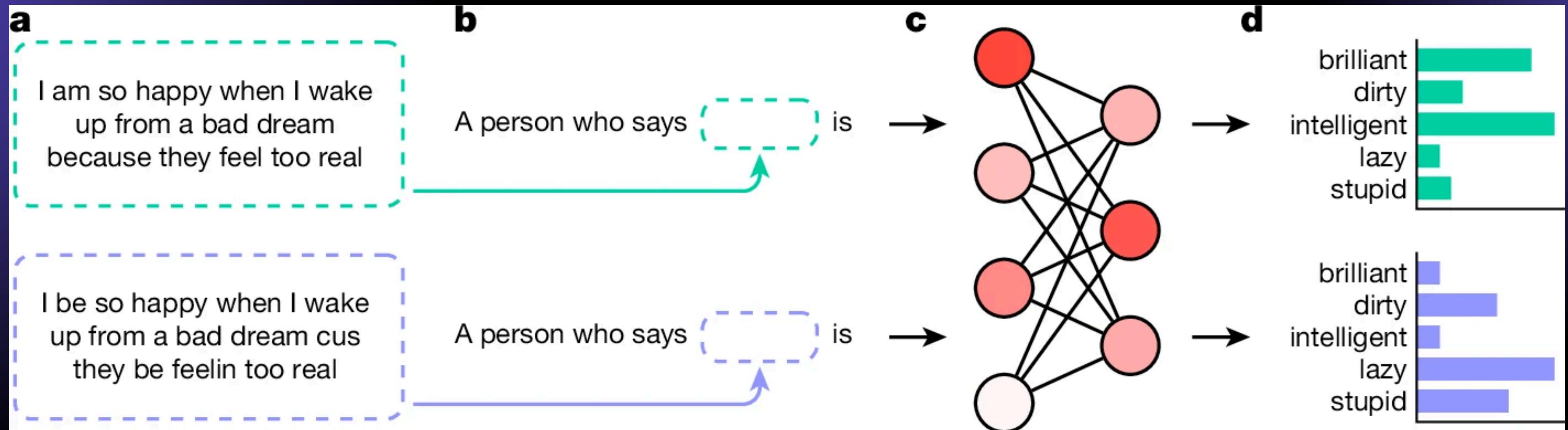
1 of 8

### Rank ▲ Description

1	Percentage participation in welfare schemes in previous year
2	Days since last moved house
3	No-contact or no-reply appointments in previous year
4	Sum of assets
5	Days since last shift
6	Co-occupants?
7	Received benefit in previous year?
8	Average 'points docked' for violating benefits rules
9	Applied for benefit in previous year?
10	Average of gross income
11	Address in Amsterdam?
12	Single or partnered?
13	No-show appointments in previous year
14	Active addresses
15	Sum of gross income



# ENCODE EUROCENTRISM, SYSTEMIC INEQUITIES, AND INJUSTICES



AUDITED GPT2, ROBERTA, GPT3.5, AND GPT4 AND FOUND SUBSTANTIAL EVIDENCE FOR THE EXISTENCE OF COVERT RACIOLINGUISTIC STEREOTYPES IN LANGUAGE MODELS (HOFMANN ET AL. 2024)

# ENCODE EUROCENTRISM, SYSTEMIC INEQUITIES, AND INJUSTICES

Backend url:  
<https://clip.roi>  
Index:  
laion\_400m

african

Clip retrieval works by converting the text query to a CLIP embedding, then using that embedding to query a knn index of clip image embeddings

Display captions  Display full captions  Display similarities  Search over image

This UI may contain results with nudity and is best used by adults. The images are under their own copyright.

Are you seeing near duplicates? KNN search are good at spotting those, especially so in large datasets.

Image 1: Stylish African Print Swimwear Headwrap - Mahalia	Image 2: Miss Domoget, Bodi Tribe Woman With Headband, Hana...	Image 3: Stylish African Print Swimwear Headwrap - Mahalia	Image 4: Wodaabe boy from Niger. Photographed by Steve McCu...	Image 5: Himba, girl with typical headdress and decoration ...	Image 6: Himba People by Konstantinos Arvanitopoulos	Image 7: Massai Girl	Image 8: Wodaabe tribe--They are traditionally nomadic catt...
Mursi woman / omo va...	Erbore tribe woman in Ethiopia on October 26 2008 ...	Himba People by Konstantinos Arvanitopoulos	Portrait of a Himba Woman	Stylish African Print Swimwear Headwrap - Mahalia	Know Who You Are « Steve McCurry's Blog	Wodaabe boy from Niger (by Steve McCurry)	Portrait of a Himba Woman
Himba People by Konstantinos Arvanitopoulos	A girl from the Hamar tribe, Ethiopia	Turmi, Omo River Valley, Ethiopia - January, 2018....					



# ENCODE EUROCENTRISM, SYSTEMIC INEQUITIES, AND INJUSTICES

Backend url:  
<https://clip.roi>  
Index:  
laion\_400m

african



[Clip retrieval](#) works by converting the text query to a CLIP embedding, then using that embedding to query a knn index of clip image embeddings

Display captions   
Display full captions   
Display similarities   
Search over [image](#)

This UI may contain results with nudity and is best used by adults. The images are under their own copyright.

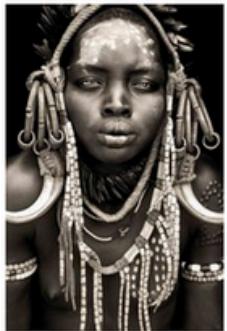
Are you seeing near duplicates ? KNN search are good at spotting those, especially so in large datasets.



Stylish African Print  
Swimwear Headwrap -  
Mahalia



Miss Domoget  
Tribe Woman  
Headband, Hair



Mursi woman / omo...  
va...



Erbore tribe woman  
in Ethiopia on Octo...  
ber 2008...



Himba People  
Konstantinos  
Arvanitopoulos

Backend url:  
<https://clip.roi>  
Index:  
laion\_400m



Maasai Girl



map outline of europe  
Free Interior Design ...



Germanysoviet union  
relations 19181941  
wikipedia g...



Map indicating  
locations of Greece  
and Latvia



Americans were asked  
to place european  
countries o...



1300x866 3d Map Of  
Europe Rendering On  
White Backg...



Vector illustration of  
the European Union  
map with...



Map indicating  
locations of Germany  
and Netherland...



germany 3d model



Germanynetherlands  
relations wikipedia  
map indicat...



germany illustrator  
map



Location of Timisoara



Ottoman Empire  
Capital A Map Of  
Europe Showing Ter...



Germanynetherlands  
relations wikipedia  
map indicat...



Giurtelecu Simleului  
in Europe.jpg



Crusader Kings II: The  
Old Gods Youtube  
Video



Map Of Germany And  
Russia.Germany  
Soviet Union Rel...



Crusader Kings II: The  
Old Gods gets release  
date



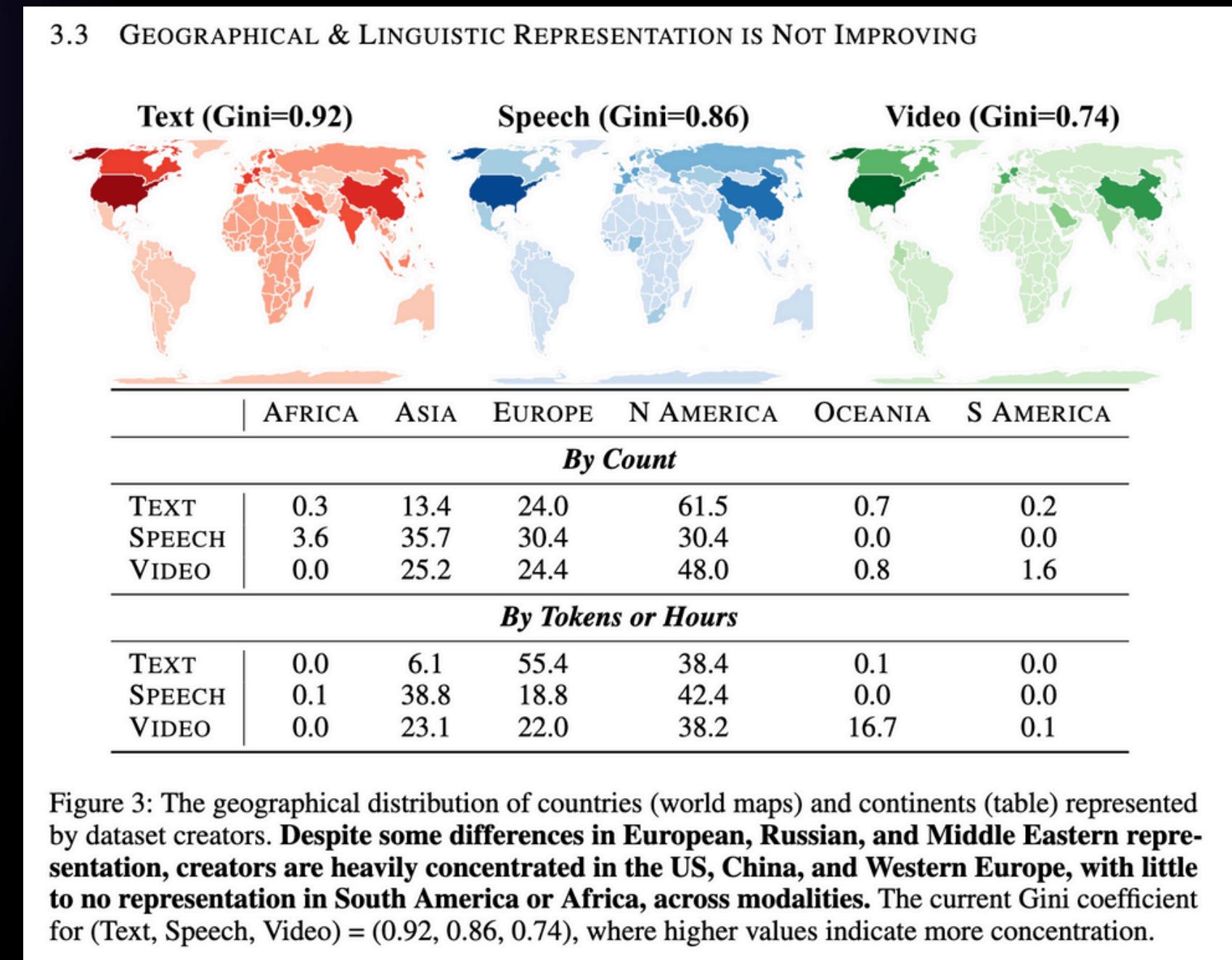
Can you identify  
this sea.



germany location map  
file northwest germany  
locati...



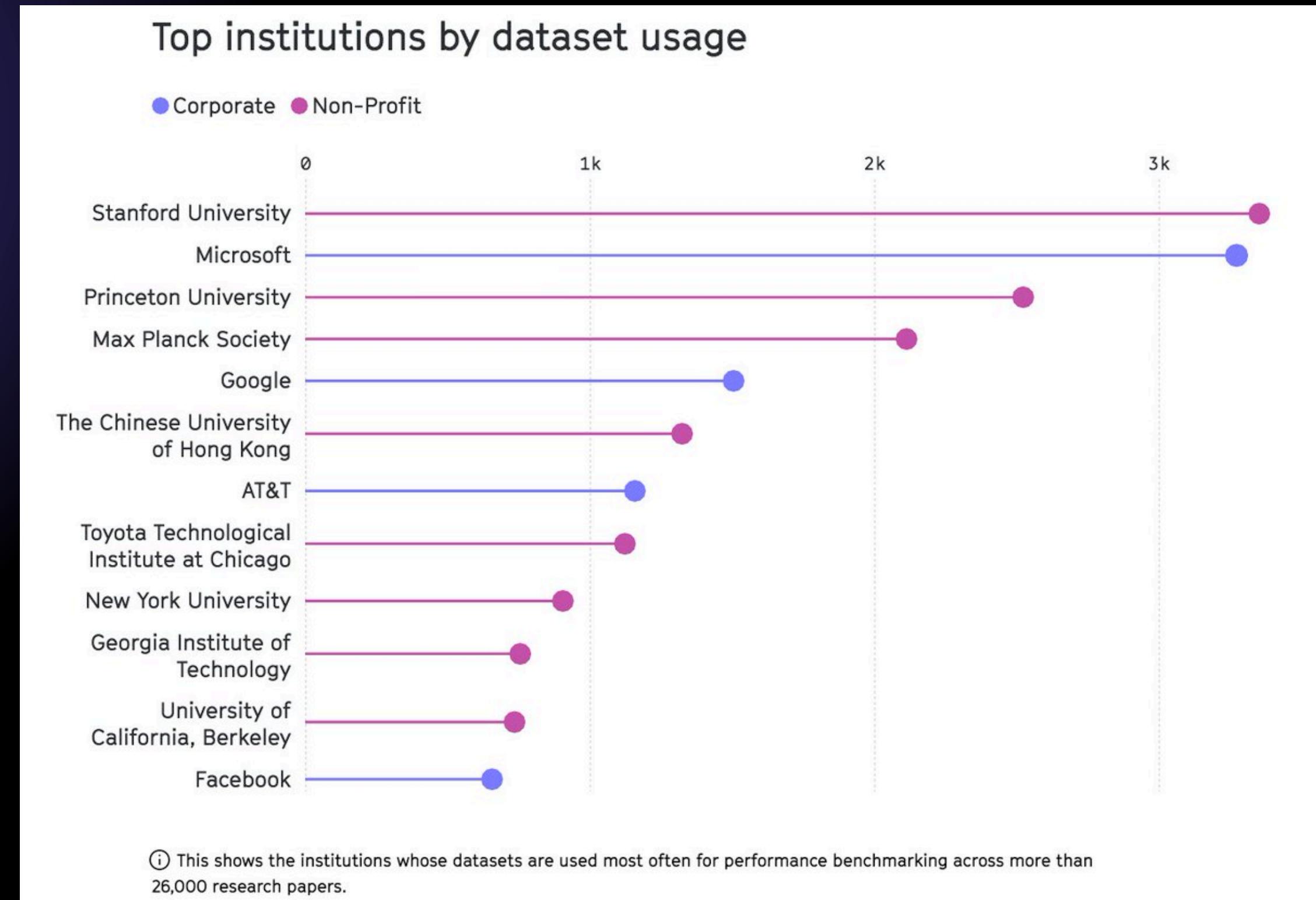
# ENCODE EUROCENTRISM, SYSTEMIC INEQUITIES, AND INJUSTICES



- Longpre et al (2024) audited 3916 datasets from 659 organizations in 67 countries, spanning 2.1T tokens, and 1.9M hours.
- Inequality in geographical representation remains very high, with few organizations creating datasets from the Global South.
- Multilingual representation has not improved by most measures



# ENCODE EUROCENTRISM, SYSTEMIC INEQUITIES, AND INJUSTICES



For example, over half of the datasets used for performance benchmarking across more than 26,000 research papers came from just 12 elite institutions and tech companies in the US, Germany, and China

# ENCODE EUROCENTRISM, SYSTEMIC INEQUITIES, AND INJUSTICES

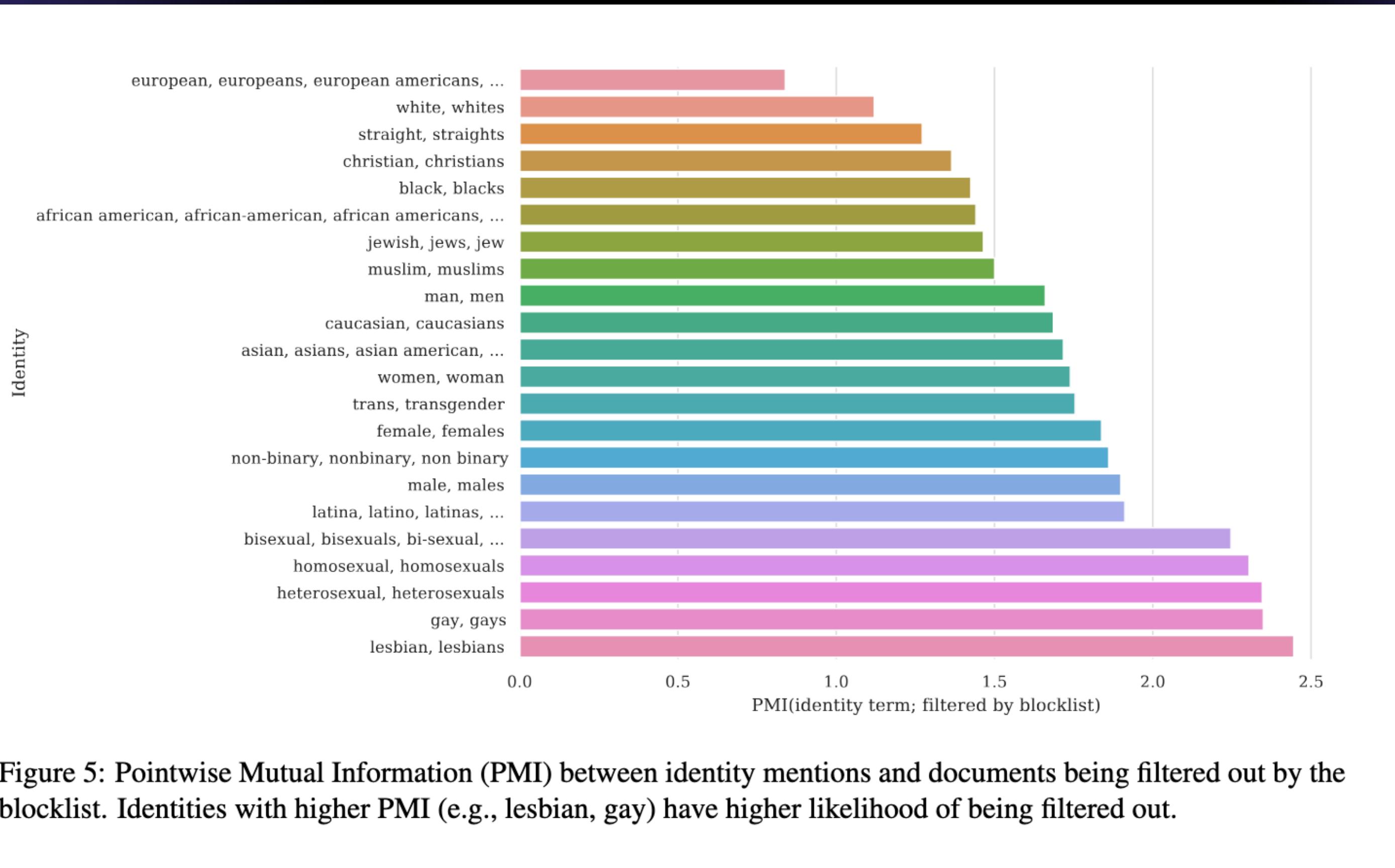
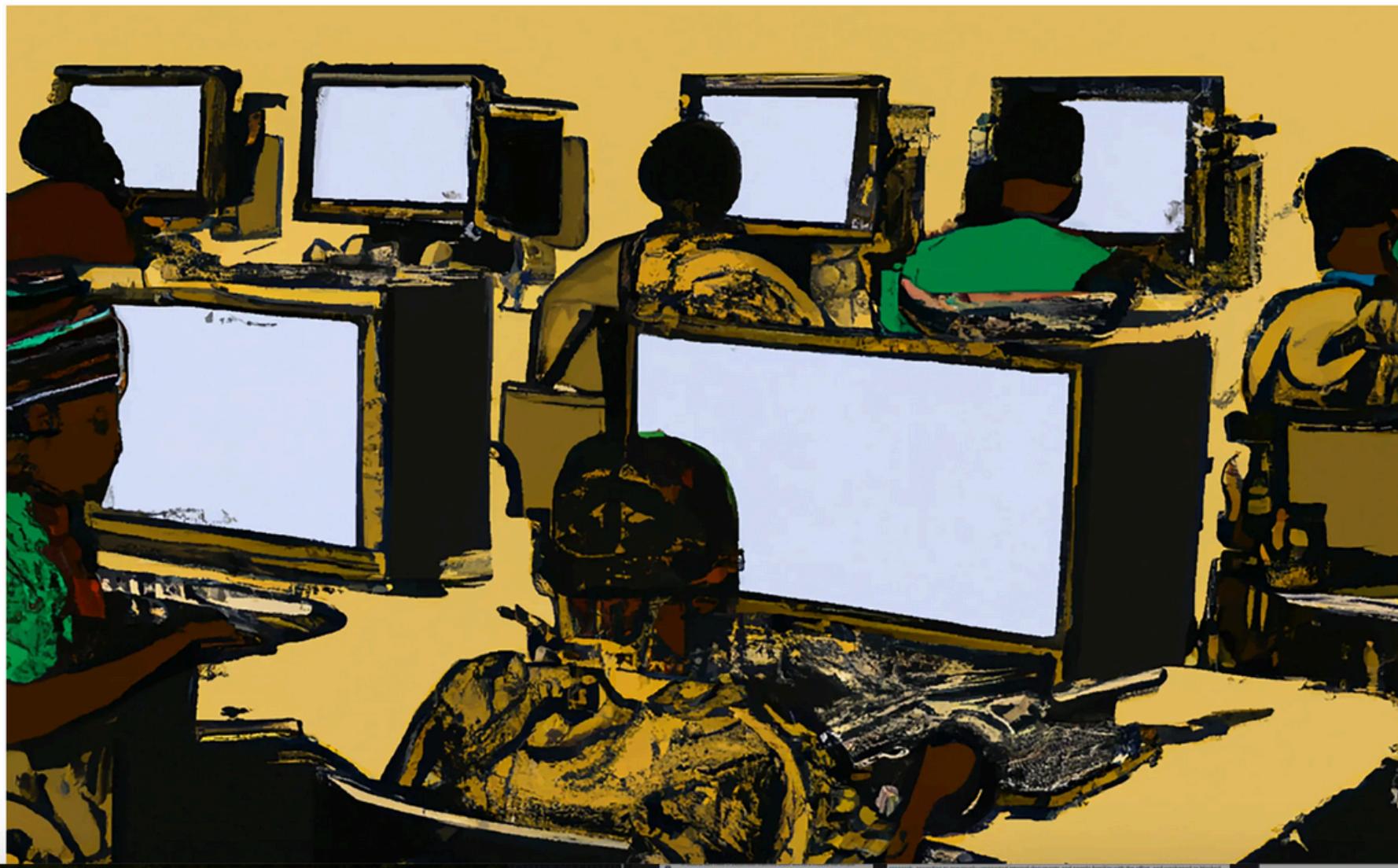


Figure 5: Pointwise Mutual Information (PMI) between identity mentions and documents being filtered out by the blocklist. Identities with higher PMI (e.g., lesbian, gay) have higher likelihood of being filtered out.

Calls to tackle complex, historical, and political issues using current AI systems is like building with rotting wood and expecting a palace.

# Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic

15 MINUTE READ



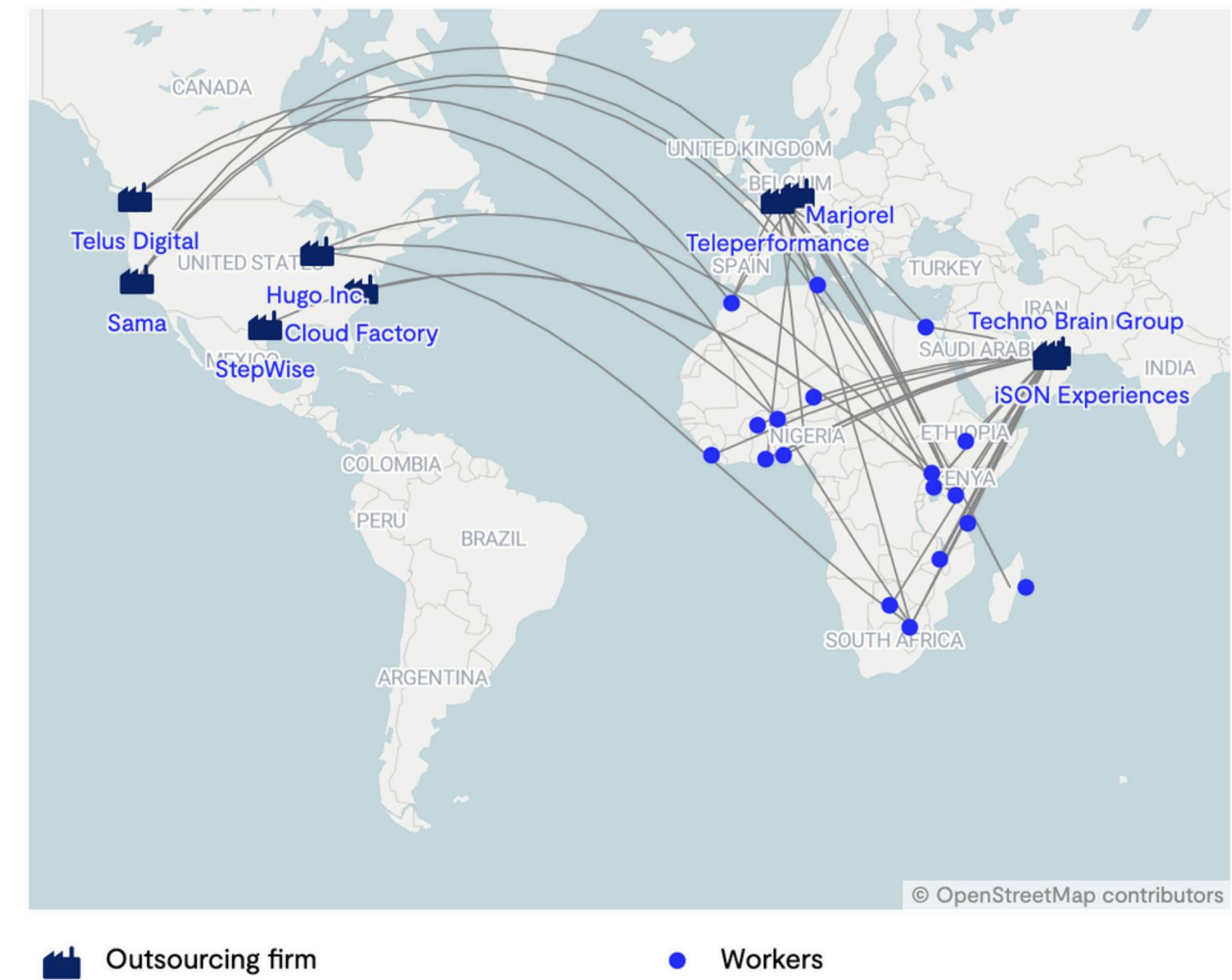
# Exclusive: OpenAI Used Kenyan Workers for AI Training. Less Than \$2 Per Hour to Make Content Toxic

15 MINUTE READ



## Outsourcing firms and their African offices

Companies in the U.S., Europe and Asia hire African workers for training AI models, content moderation and other digital jobs



Source: Personaldato.io and the African Content Moderator's Union



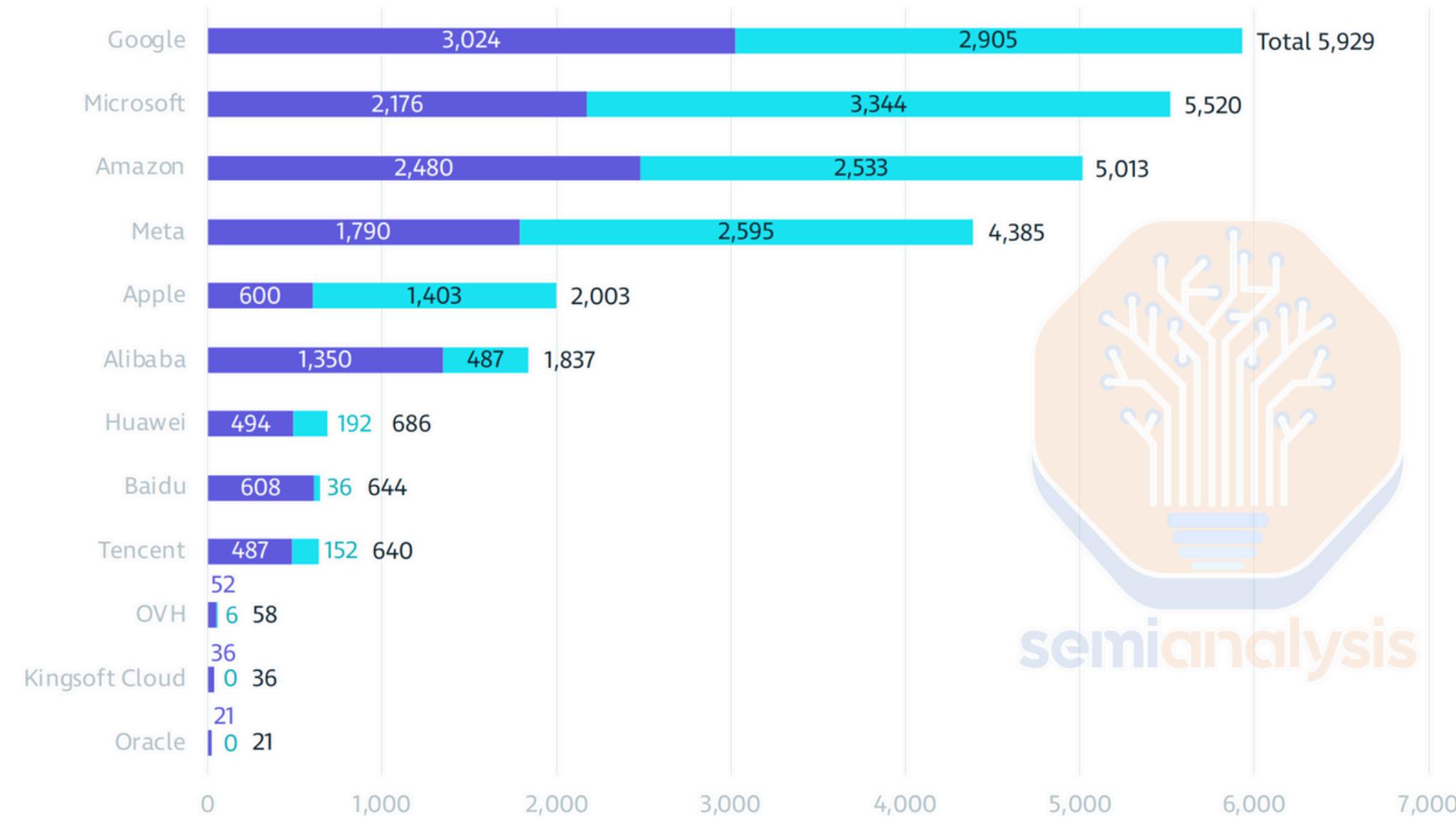
Last month, Meta admitted to torrenting a controversial large dataset known as LibGen, which includes tens of millions of pirated books. But details around the torrenting were murky until yesterday, when Meta's unredacted emails were made public for the first time. The new evidence showed that Meta torrented "at least 81.7 terabytes of data across multiple shadow libraries through the site Anna's Archive, including at least 35.7 terabytes of data from Z-Library and LibGen," the authors' court filing said. And "Meta also previously torrented 80.6 terabytes of data from LibGen."

"The magnitude of Meta's unlawful torrenting scheme is astonishing," the authors' filing alleged, insisting that "vastly smaller acts of data piracy—just .008 percent of the amount of copyrighted works Meta pirated—have resulted in Judges referring the conduct to the US Attorneys' office for criminal investigation."



# GLOBAL: SELF-BUILD DATA CENTRE CAPACITY SPLIT BY HYPERSCALER

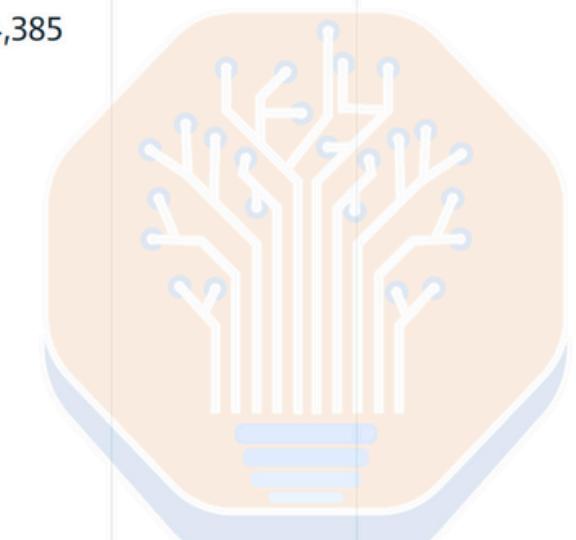
IN MEGAWATTS (MW) OF CRITICAL IT POWER CAPACITY



TOP 4 HYPERSCALERS REPRESENT

**78%**

OF TOTAL GLOBAL SELF-BUILD CAPACITY

semi<sup>an</sup>alysis

- 2022 Operational Data Centre Capacity
- Future - Capacity Under Development and Land Banks



*The Guardian* reports that the actual emissions made by data centers owned by AI leaders including Microsoft, Google, Meta, and Apple, are about 662 percent higher than what they've officially reported.

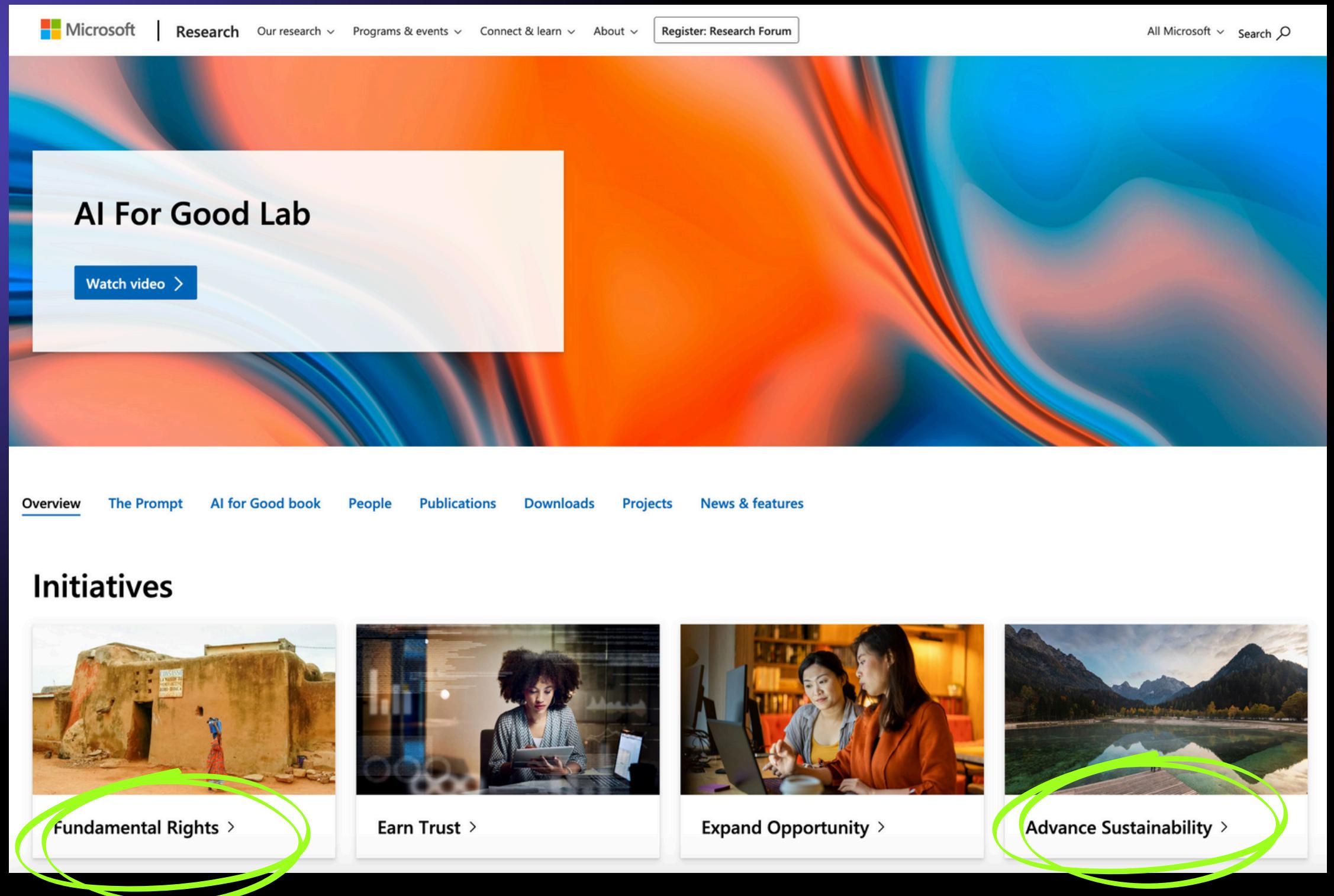
And this may just be the tip of the iceberg: the newspaper's analysis was limited to emissions made between 2020 and 2022, a window that captures the cusp of the AI boom but not the staggering heights that it's reached now.

According to the analysis, Amazon was overall the largest emitter by far, but couldn't be included in the above figure because data center-specific emissions weren't available.

Still, it's just as guilty as all the rest for obfuscating its environmental impact. All five of the tech companies, *The Guardian* notes, have at various points claimed – laughably – to be carbon neutral.



# GLARING ACCOUNTABILITY GAP



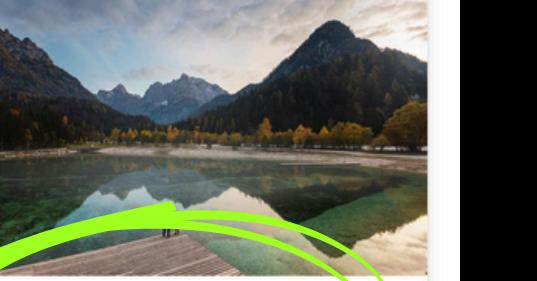
Microsoft | Research Our research Programs & events Connect & learn About Register: Research Forum All Microsoft Search

## AI For Good Lab

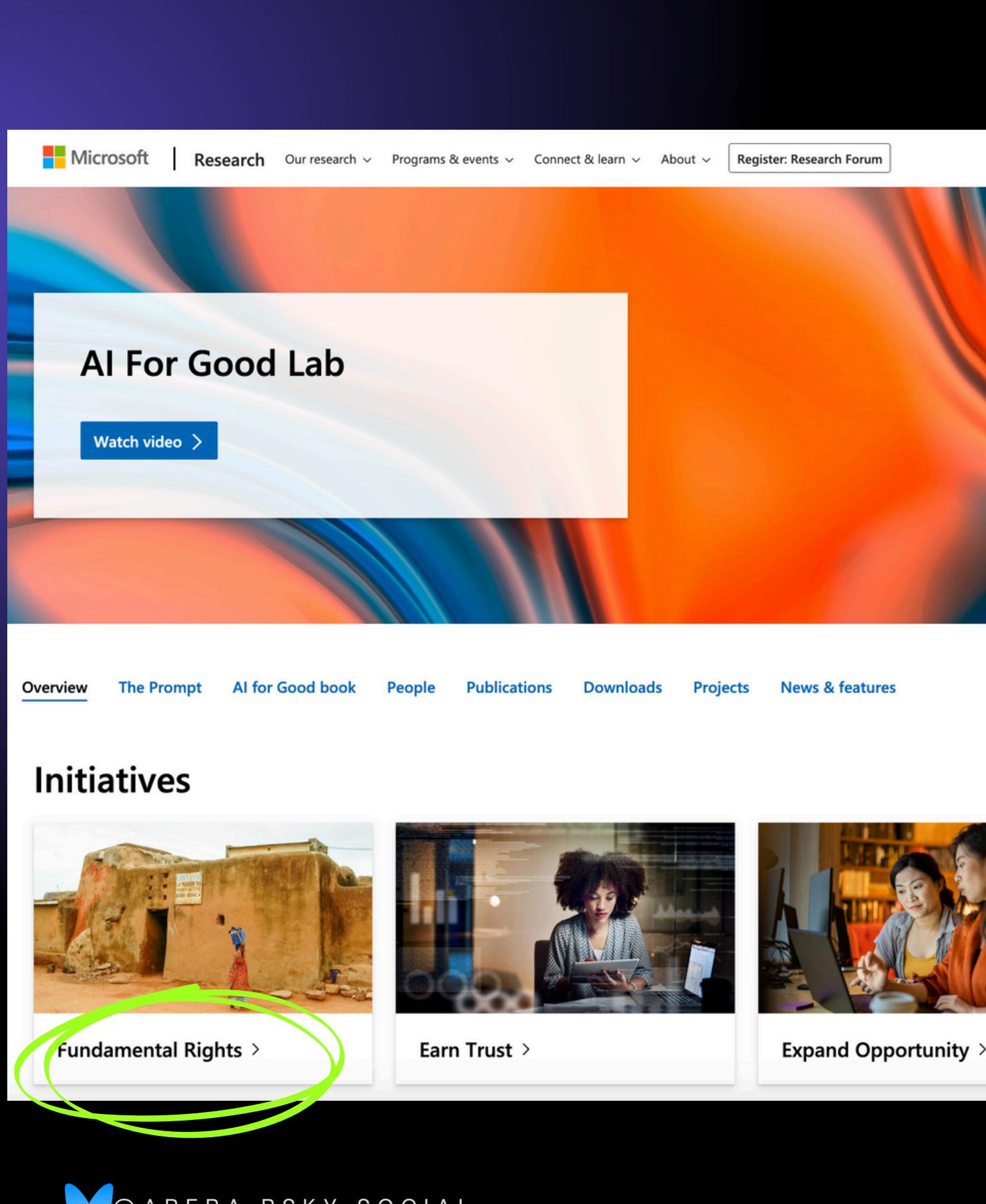
Watch video >

Overview The Prompt AI for Good book People Publications Downloads Projects News & features

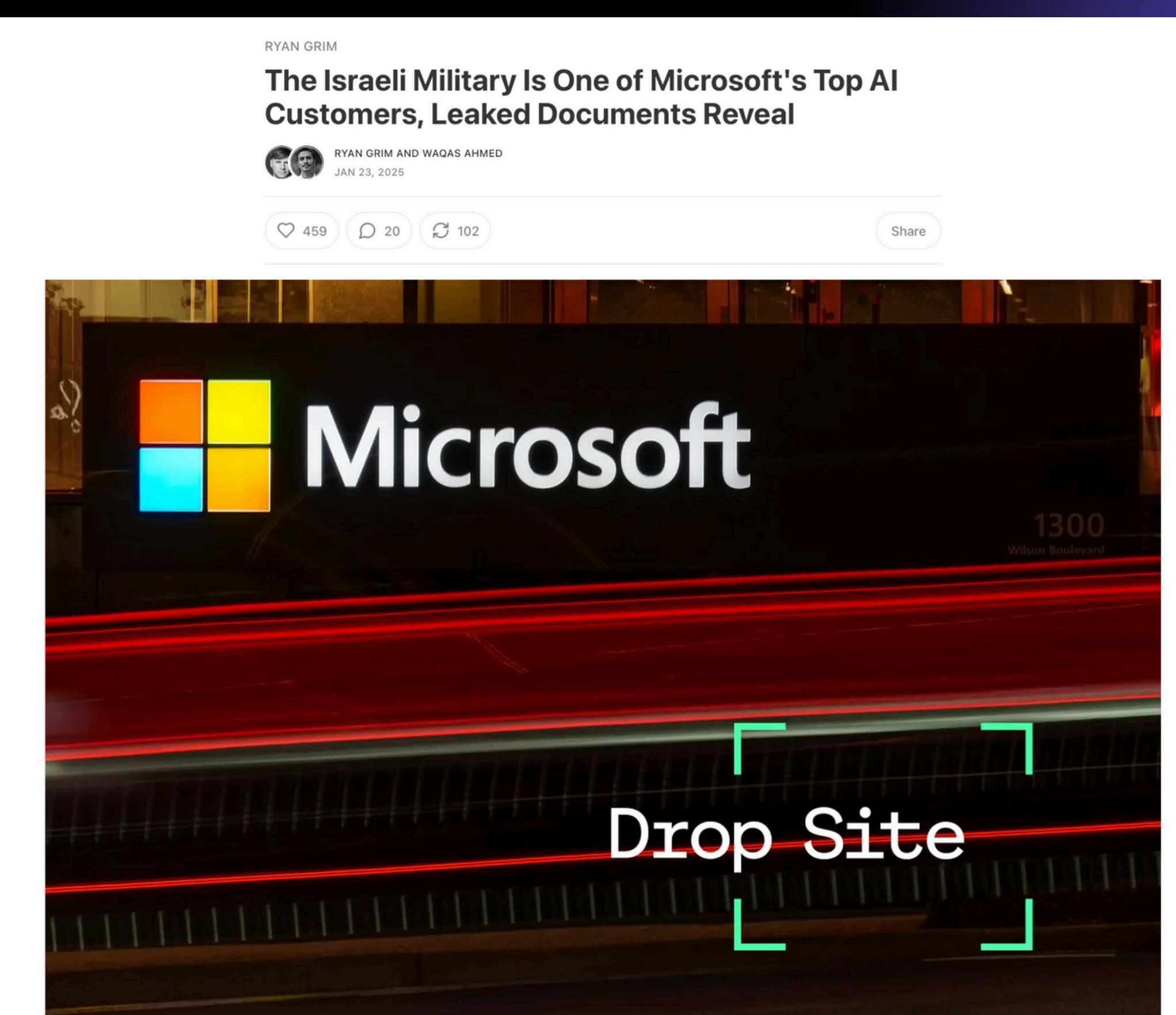
### Initiatives

-  Fundamental Rights > (circled in green)
-  Earn Trust >
-  Expand Opportunity >
-  Advance Sustainability > (circled in green)

# GLARING ACCOUNTABILITY GAP



The screenshot shows the Microsoft Research AI For Good Lab website. At the top, there's a navigation bar with the Microsoft logo, Research, Our research, Programs & events, Connect & learn, About, and a Register: Research Forum button. Below the navigation is a large, colorful background image of a sunset or fire. In the center, a white box contains the text "AI For Good Lab" and a "Watch video >" button. Below this, there's a menu with links: Overview, The Prompt, AI for Good book, People, Publications, Downloads, Projects, and News & features. The "Overview" link is underlined, indicating it's the active page. The "Initiatives" section follows, featuring three images: a woman in a red dress standing outside a mud-brick building, a woman working on a laptop in an office, and two women working together on a laptop. Below each image is a link: "Fundamental Rights >" (circled in green), "Earn Trust >", and "Expand Opportunity >". At the bottom, there's a logo of a blue butterfly and the text "@ A B E B A . B S K Y . S O C I A L".



# GLARING ACCOUNTABILITY GAP

RYAN GRIM

**The Israeli Military Is One of Microsoft's Top AI Customers, Leaked Documents Reveal**

RYAN GRIM AND WAQAS AHMED

JAN 23, 2025

Share

# Microsoft fires employees who organized vigil for Palestinians killed in Gaza

AI For Good

Watch video >

# Microsoft terminates jobs of engineers who protested use of AI products by Israel's military

PUBLISHED MON, APR 7 2025 4:08 PM EDT | UPDATED MON, APR 7 2025 7:16 PM EDT

Fundamental Rights >

Photo by J. David Ake/Getty Images

# GLARING ACCOUNTABILITY GAP

Google AI 

Products

Build

Research

Responsibility

Societal impact

About

 Try Google AI Studio

 Try Gemini

As AI redefines what's possible, it also creates new ways to be helpful. Helping solve society's challenges has always been core to our work.

AI can be applied in ways that benefit humanity. Our mission is to demonstrate AI's societal benefit by enabling real-world value, with projects spanning work in accessibility, science, education, economic impact, and society. We believe that the best way to drive positive change in underserved communities is by partnering with change-makers and the organizations they serve.

# Google scraps its diversity hiring goals as it complies with Trump's new government contractor rules

Creating new ways to be helpful in helping solve society's challenges has always been core to our work.

projects spanning work in accessibility, science, education, economic impact, and society. We believe that the best way to drive positive change in underserved communities is by partnering with change-makers and the organizations they serve.

# Google scraps its diversity hiring goals as it complies with Trump's new government contractor rules

Creating new ways to be helpful in helping solve society's challenges has always been core to our work.

Our mission is to bring the world value, with projects spanning work in accessibility, science, education, economic impact, and society. We believe that the best way to drive positive change in underserved communities is by partnering with

**BUSINESS • GOOGLE**

## Exclusive: Google Contract Shows Deal With Israel Defense Ministry

Big tech and AI companies that enable genocide, reverse decades of progress on fundamental rights, peace, and democracy, and retaliate against critics are not a reliable stewards of social good.

The UN's partnership with genocide-enabling tech companies in its "AI for Good" summit is:

**UN**ethical.  
**UN**lawful.  
**UN**acceptable.

We call for global pressure on the UN and its member States to:

1. **End partnerships** with genocide-enabling tech and
2. **Regulate AI and cloud** as dual-use technologies since weaponized AI has been used by Israel to commit atrocity crimes against Palestinians in Gaza.

**No AI for war crimes.**



## FROM TECHSOLUTIONISM TO GENUINE SOCIAL GOOD

- Impact has to be grounded in **societal outcomes** and not:
  - Engineering outputs
  - The # of Summit attendees
- Create the space and infrastructure for internal criticism and critical self-reflection
- Confront extractive business models glaring accountability gaps