




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

AI FOR SOCIAL GOOD: A TREND


 **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 >](#)


[Earn Trust >](#)

[Expand Opportunity >](#)

[Advance Sustainability >](#)



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

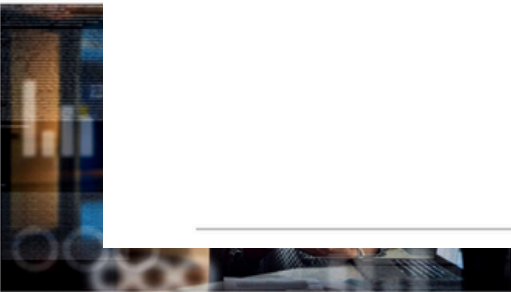
As AI redefines what's possible, it also creates new ways to be helpful. Helping solve society's challenges has always been core to our work.

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.

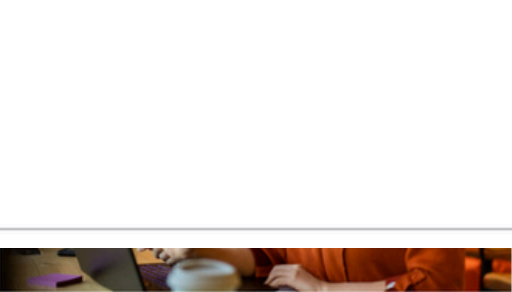
Initiatives



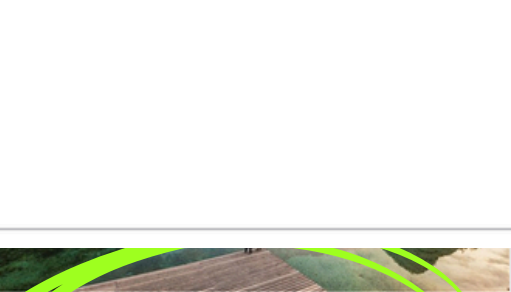
Fundamental Rights >



Earn Trust >



Expand Opportunity >



Advance Sustainability >



AI FOR SOCIAL GOOD: A TREND

United Nations University
Centre for Policy Research

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.**

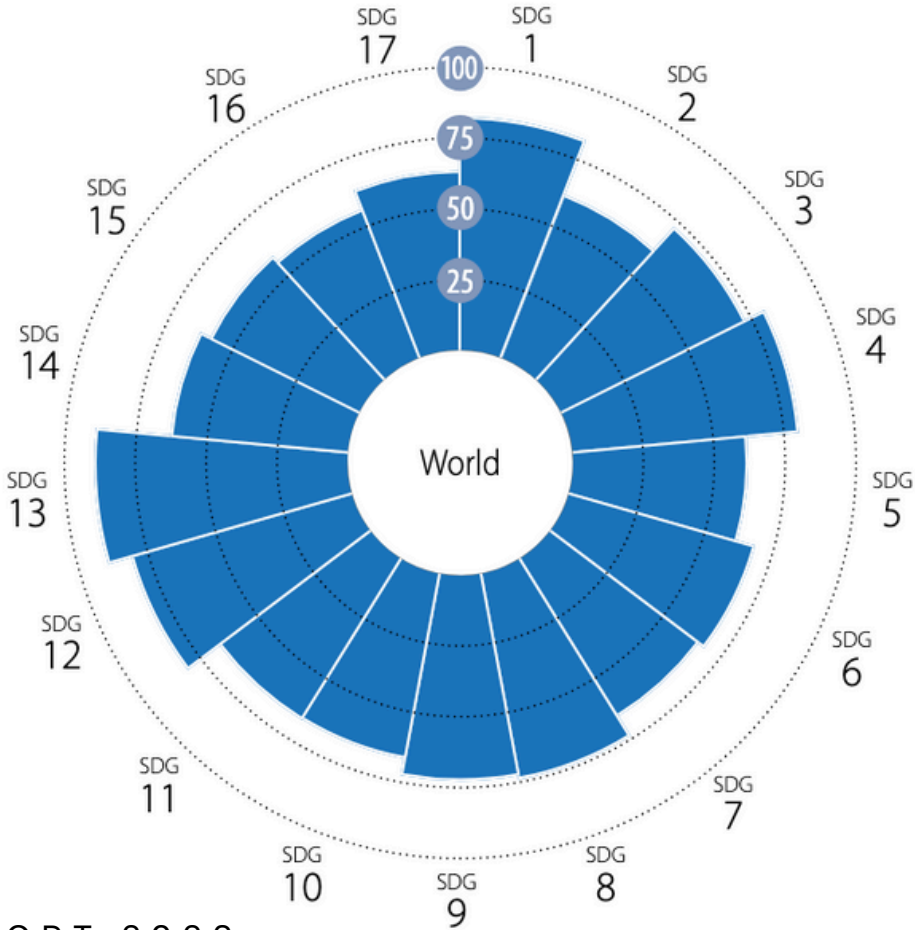
UN 2025



Overall Performance



Average Performance by SDG



SDG PROGRESS REPORT 2023

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



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

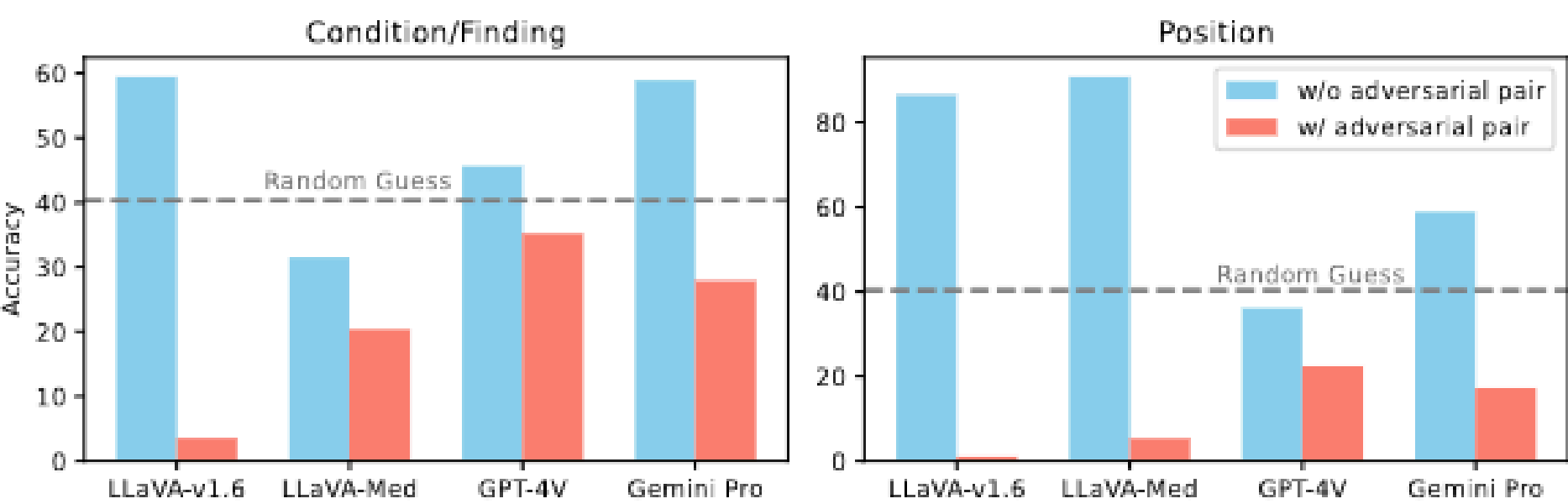


Figure 1: Accuracy of four LMMs on two types of specialized questions in medical diagnoses, with and without adversarial pairs. The significant drop in accuracy with adversarial pairs highlights the models' unreliability in handling medical diagnoses.

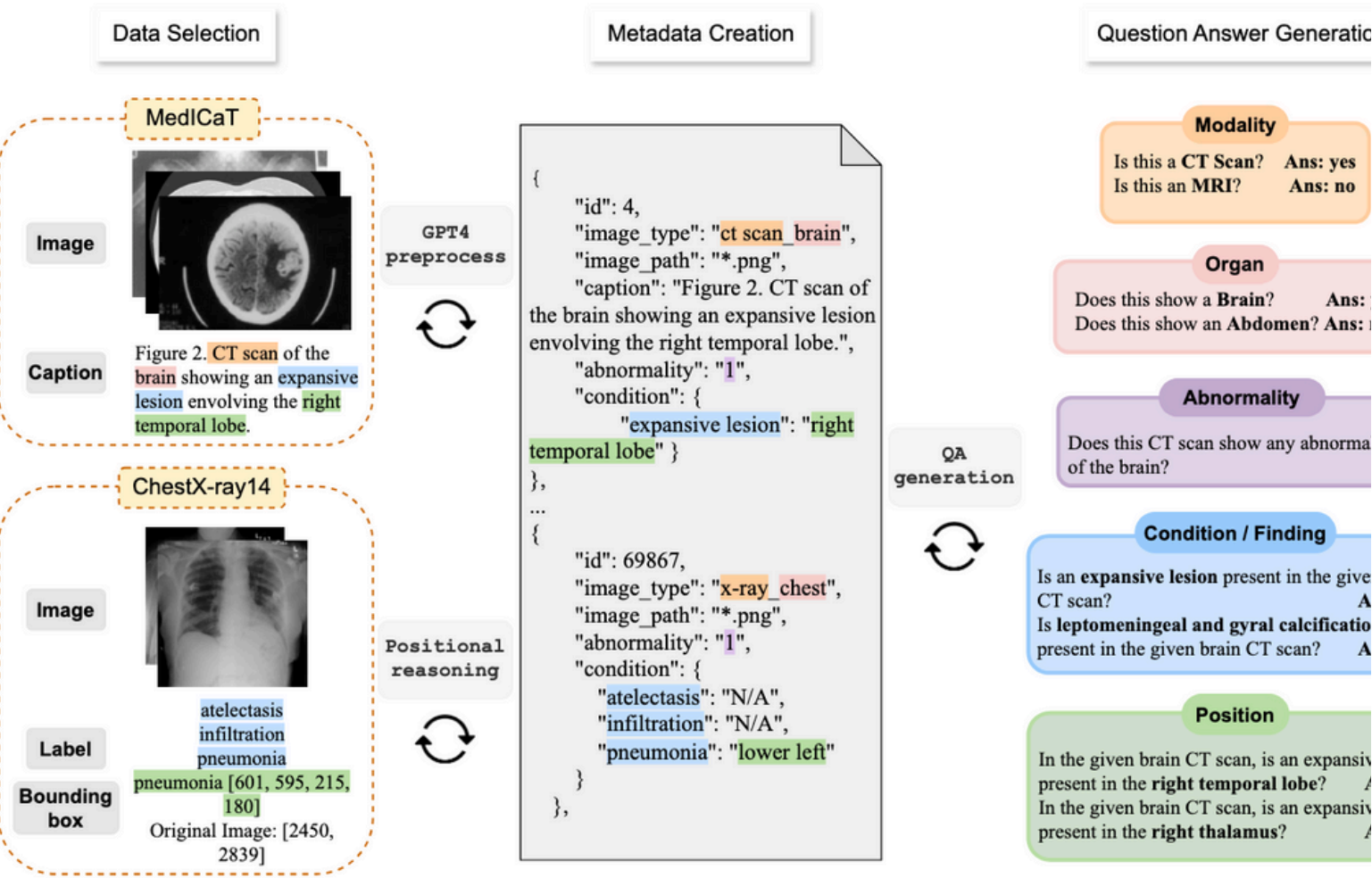


Figure 3: Flow diagram of the ProbMed data curation process. Two comprehensive biome datasets were utilized to collect source data and construct a metadata file, enabling the automated generation of high-quality question-answer pairs for the ProbMed dataset.

MAGESH ET AL., 2024

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 ³	64.16	59.16
Average ⁴	60.97	59.16
Context length ⁵	8k	8k
Model size ⁶	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
Total (out of 75)	35	61
Percentage of total	47%	81%

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
Total of documents (out of 75)	35	61
Percentage of total	47%	81%

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

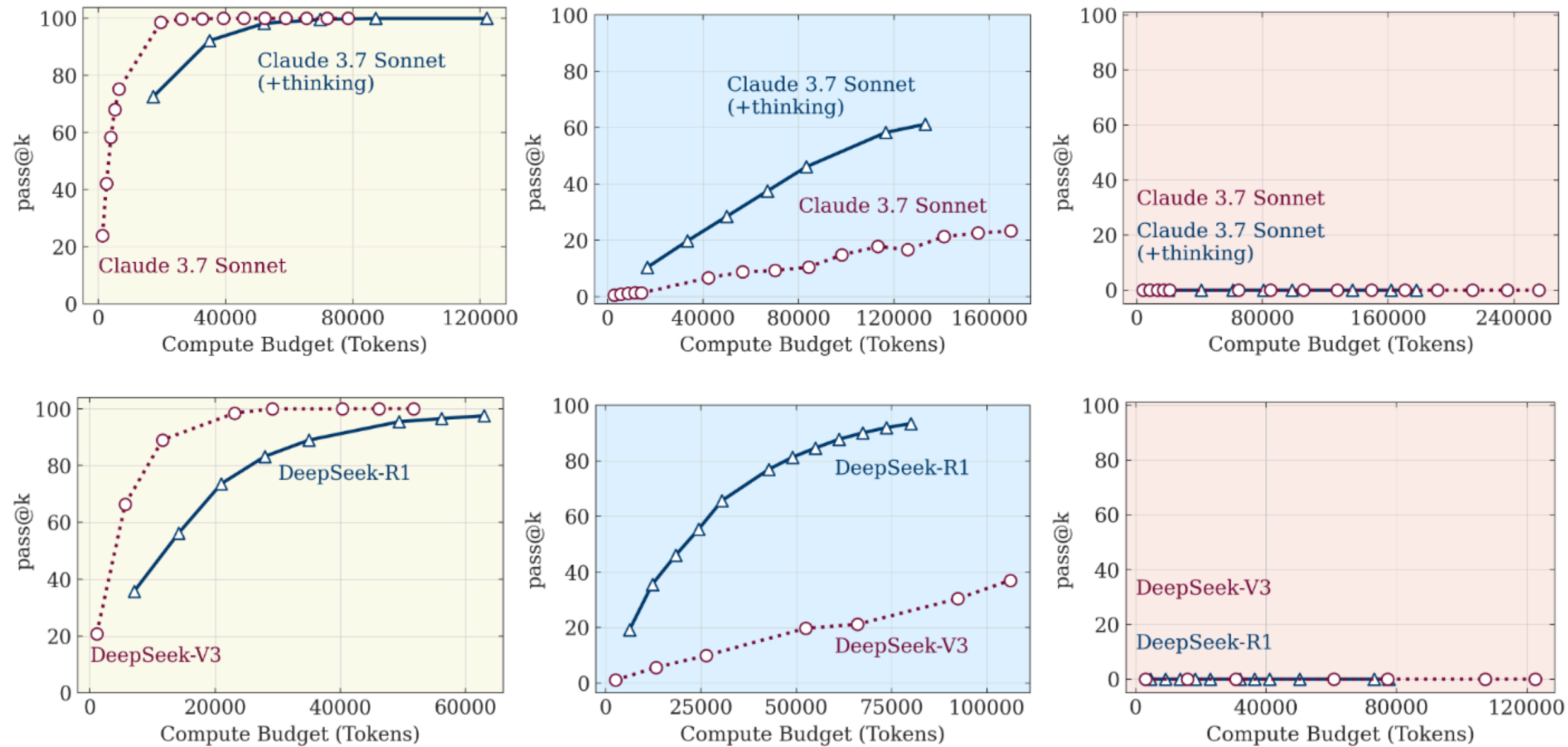
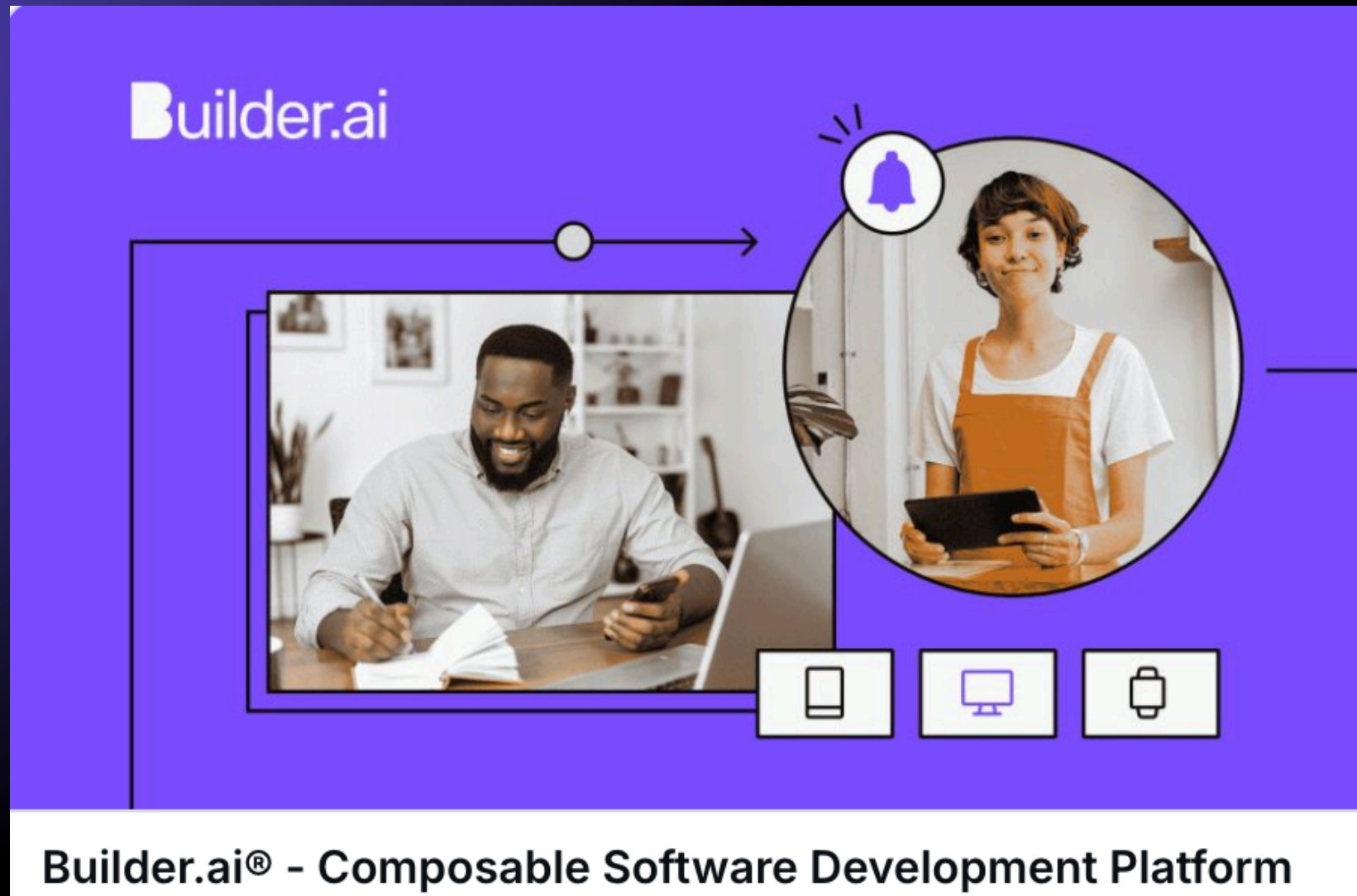


Figure 5: Pass@k performance of thinking vs. non-thinking models across equivalent compute budgets in puzzle environments of low, medium, and high complexity. Non-thinking models excel in simple problems, thinking models show advantages at medium complexity, while both approaches fail at high complexity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

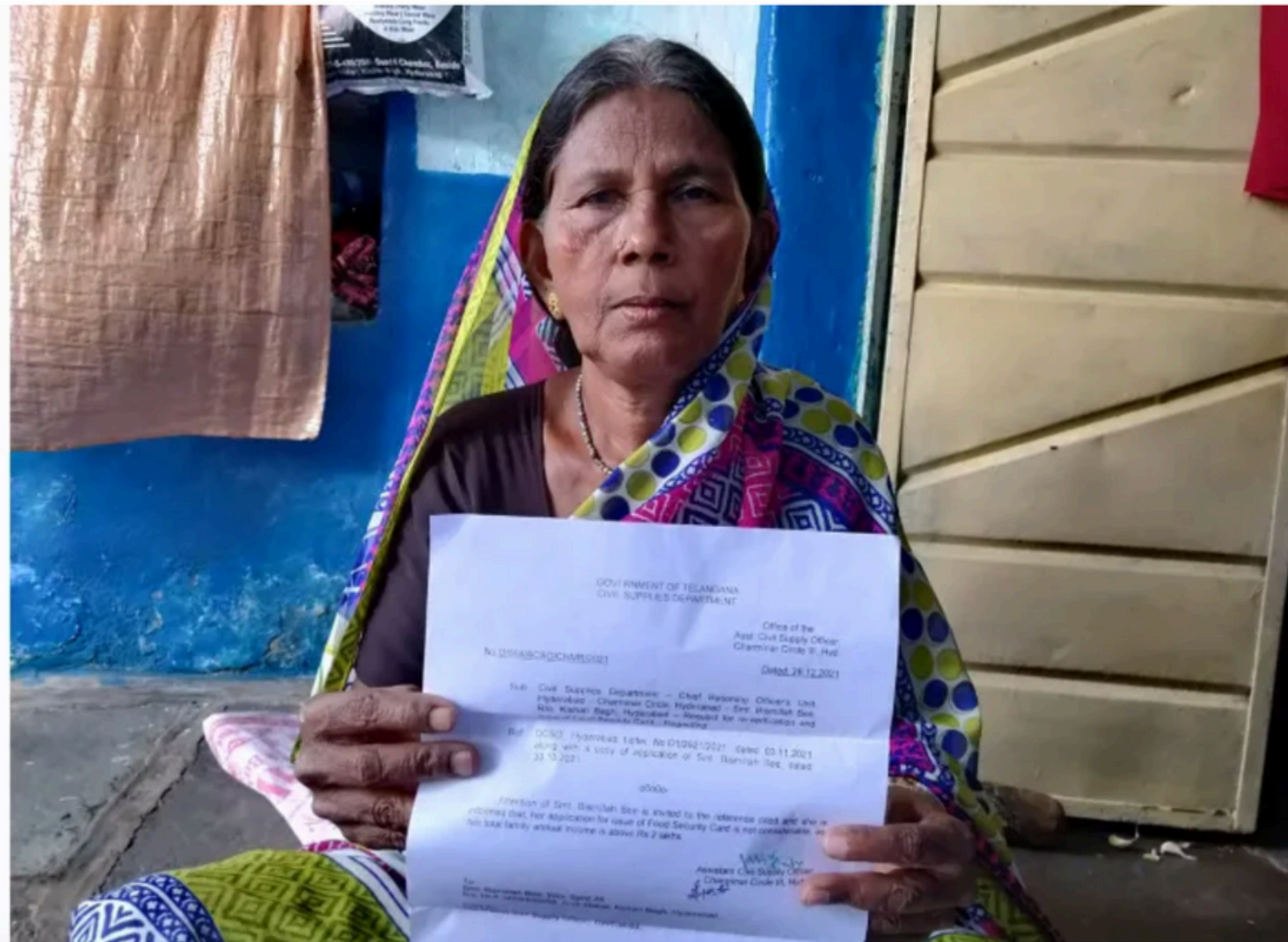
For eight years, Builder.ai marketed its "Natasha" AI system as a fully autonomous tool that could build software "as easily as ordering pizza." However, internal documents and employee accounts reviewed by Bloomberg paint a sharply different picture. Engineers in Noida and Bangalore manually coded client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

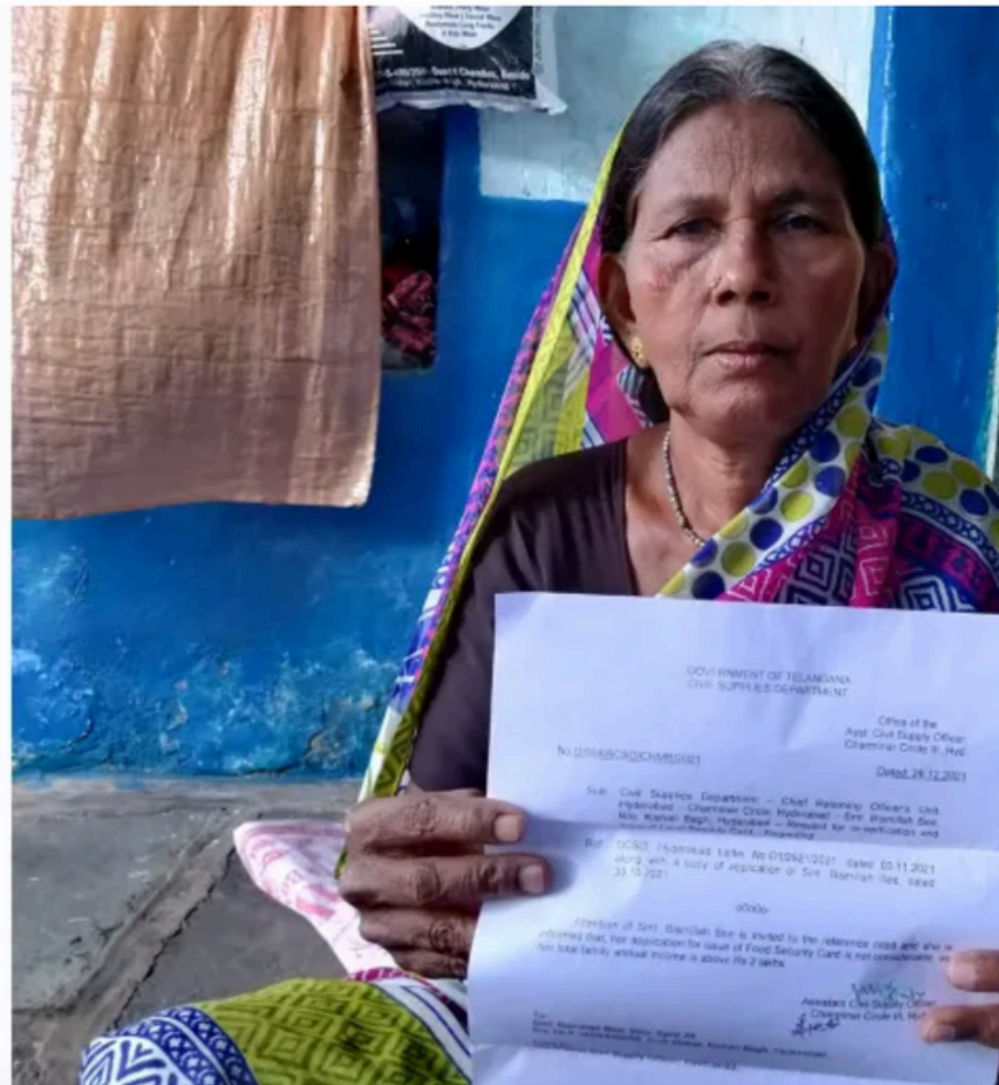
It adopted AI in welfare schemes to weed out ineligible ones, but has wrongfully removed thousands of legitimate ones.



ALJAZEERA. JAN 2024

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

It adopted AI in welfare schemes to weed out ineligible beneficiaries and wrongly removed thousands of legitimate or



ALJAZEERA. JAN 2024

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 Print

[list](#)

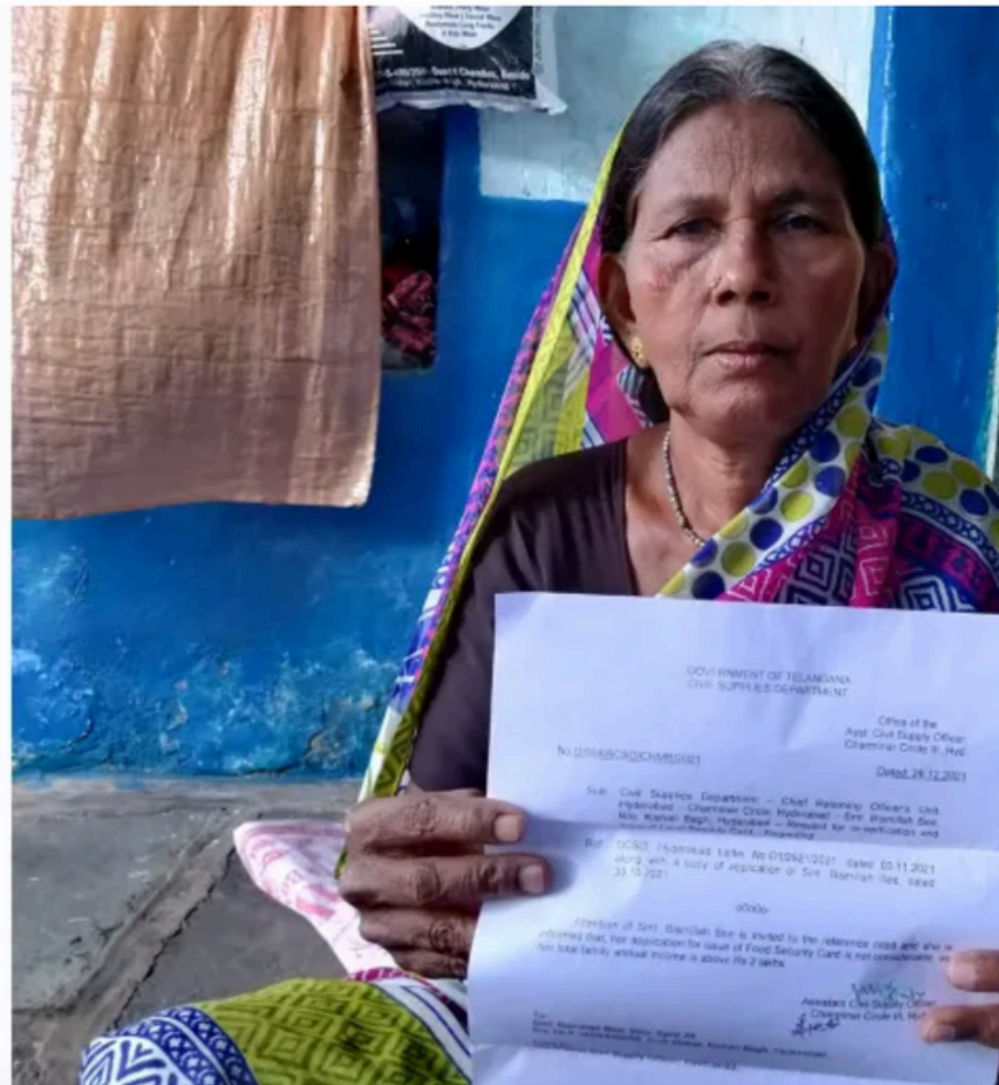


THE ASAHI SHIMBUN. MAR 2025

AI TECHNOLOGIES IN THE REAL WORLD

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

It adopted AI in welfare schemes to weed out ineligible beneficiaries and wrongly removed thousands of legitimate or



ALJAZEERA. JAN 2024

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



THE ASAHI SHIMBUN. MAR 2025

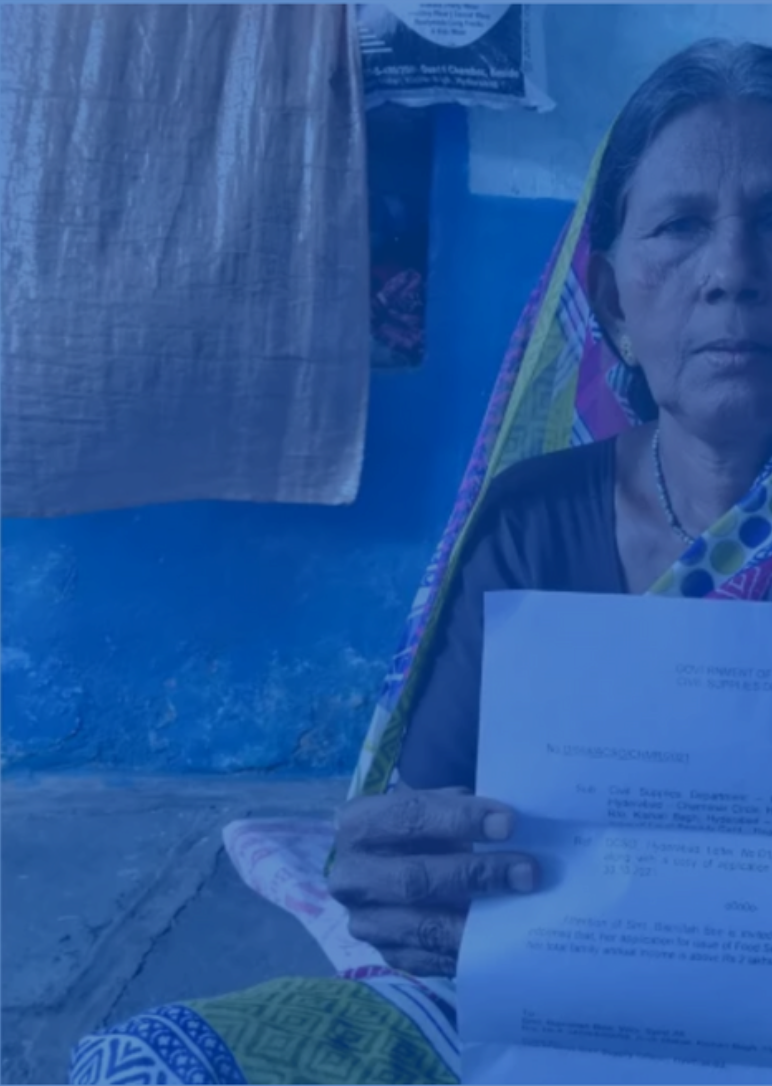
A black and white photograph of a crowded city street, likely in Amsterdam. Overlaid on the image are yellow silhouettes of people and various text elements. The text includes 'ARTIFICIAL INTELLIGENCE', 'Inside Amsterdam's high-stakes experiment to create fair welfare AI', and a list of questions related to residency and welfare. The questions are: '5.4 Sex', '5.5 Date of birth', 'Place of birth', 'Country of origin', 'Nationality', 'Do you have a residence permit?', 'Are you currently applying for a residence permit?', 'Are you unable to work?', 'Do you have a residence permit from another country?', 'Do you have a residence permit from a country in the EEA or Switzerland?', 'Do you have a residence permit from a country in the EEA or Switzerland?', 'Do you have a residence permit from a country in the EEA or Switzerland?'. The text is in a mix of English and Dutch.

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



How an algorithm denied thousands of poor in London

It adopted AI in welfare schemes to use but it has wrongly removed thousands of legal residents



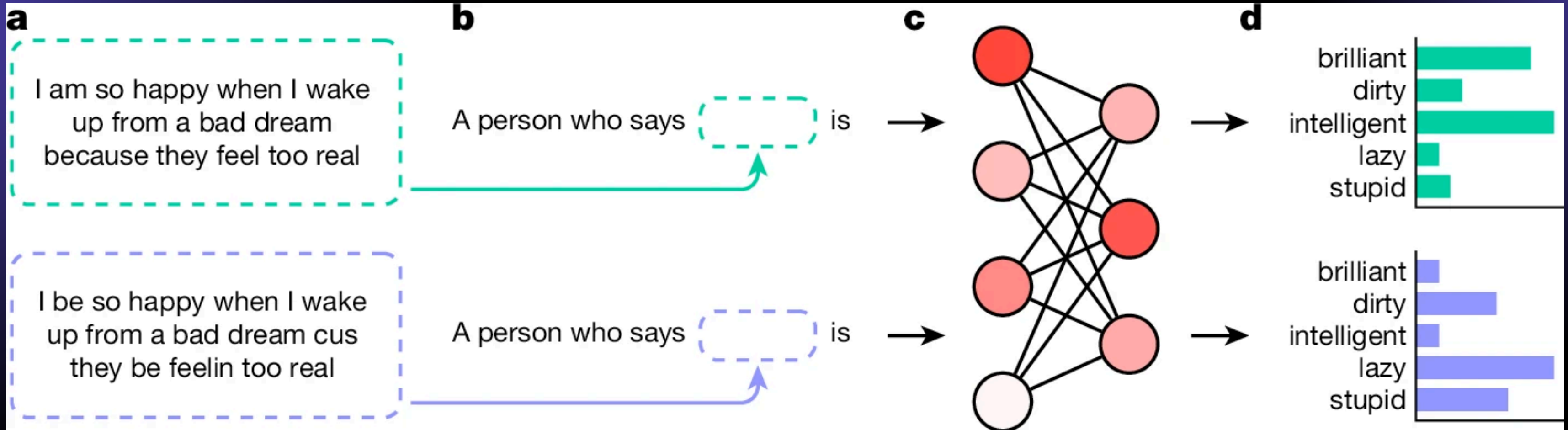
Amsterdam’s model used 15 criteria to score welfare applications.

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:


Index:

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


Display captions ☒
Display full captions ☐
Display similarities ☐
Search over

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, Bodi Tribe Woman With Headband, Hana...




Stylish African Print Swimwear Headwrap - Mahalia




Wodaabe boy from Niger. Photographed by Steve McCu...




Himba, girl with typical headdress and decoration ...




Himba People by Konstantinos Arvanitopoulos




Massai Girl




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



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

3.3 GEOGRAPHICAL & LINGUISTIC REPRESENTATION IS NOT IMPROVING

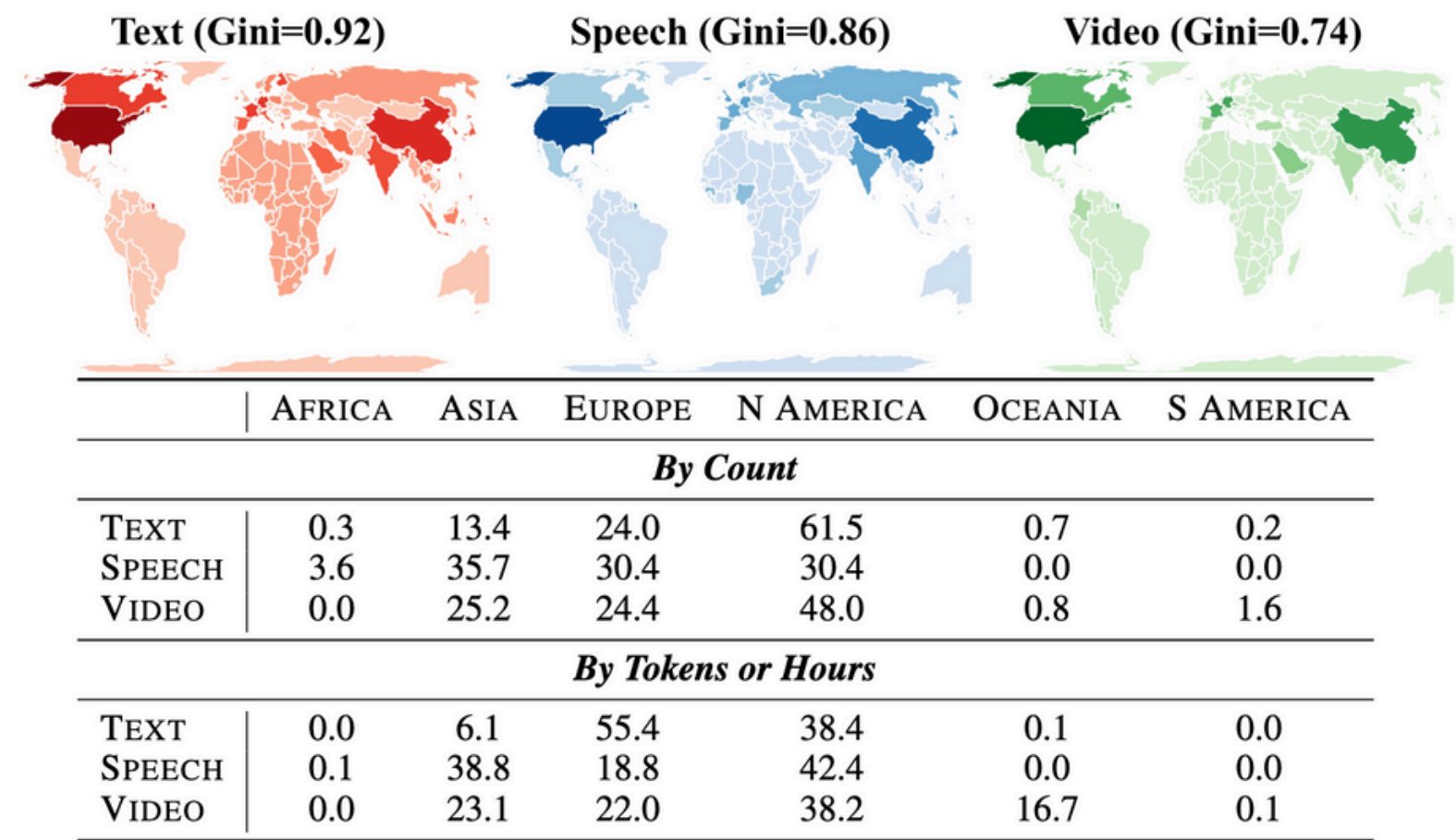
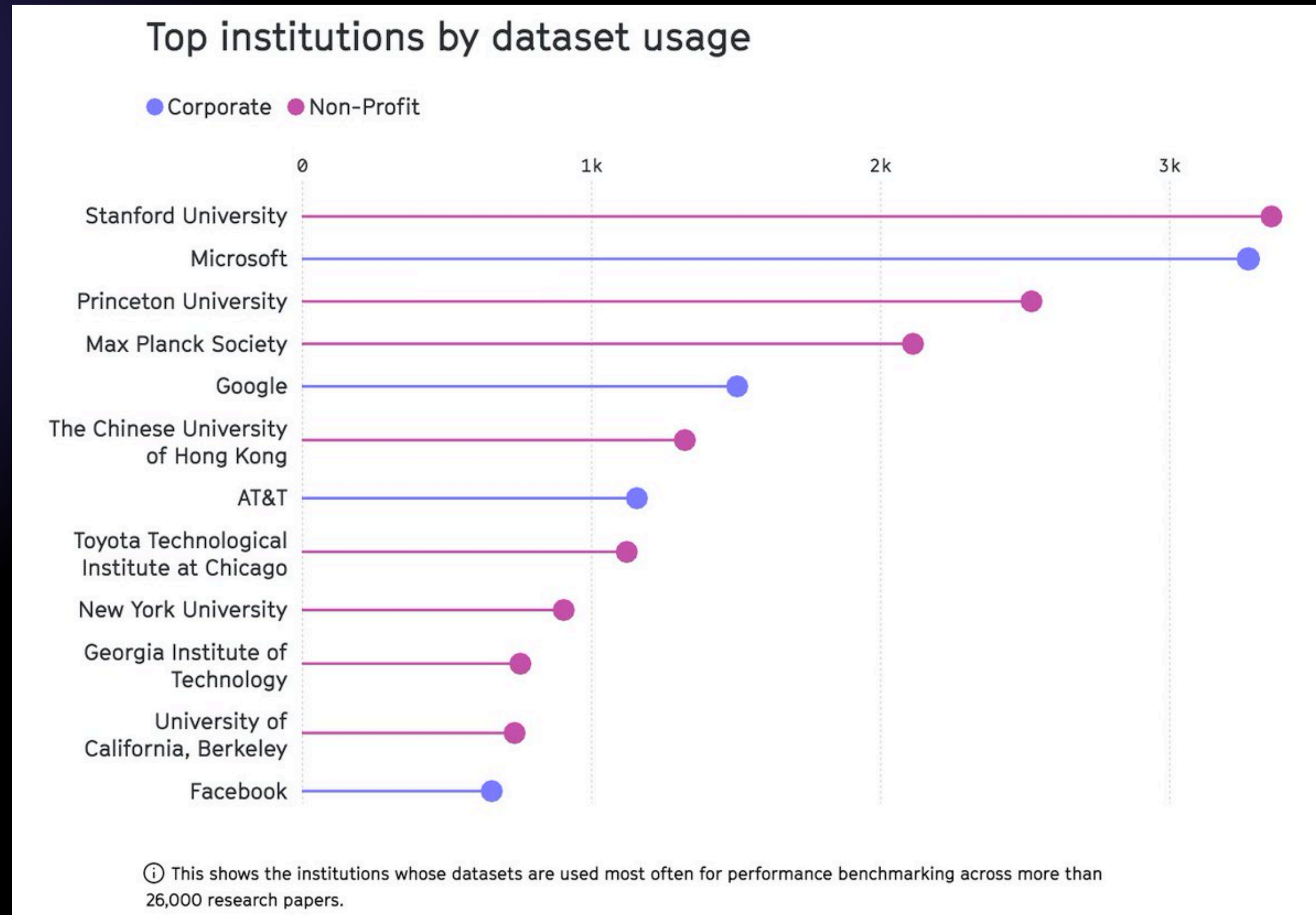


Figure 3: The geographical distribution of countries (world maps) and continents (table) represented by dataset creators. **Despite some differences in European, Russian, and Middle Eastern representation, creators are heavily concentrated in the US, China, and Western Europe, with little to no representation in South America or Africa, across modalities.** The current Gini coefficient for (Text, Speech, Video) = (0.92, 0.86, 0.74), where higher values indicate more concentration.

- 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

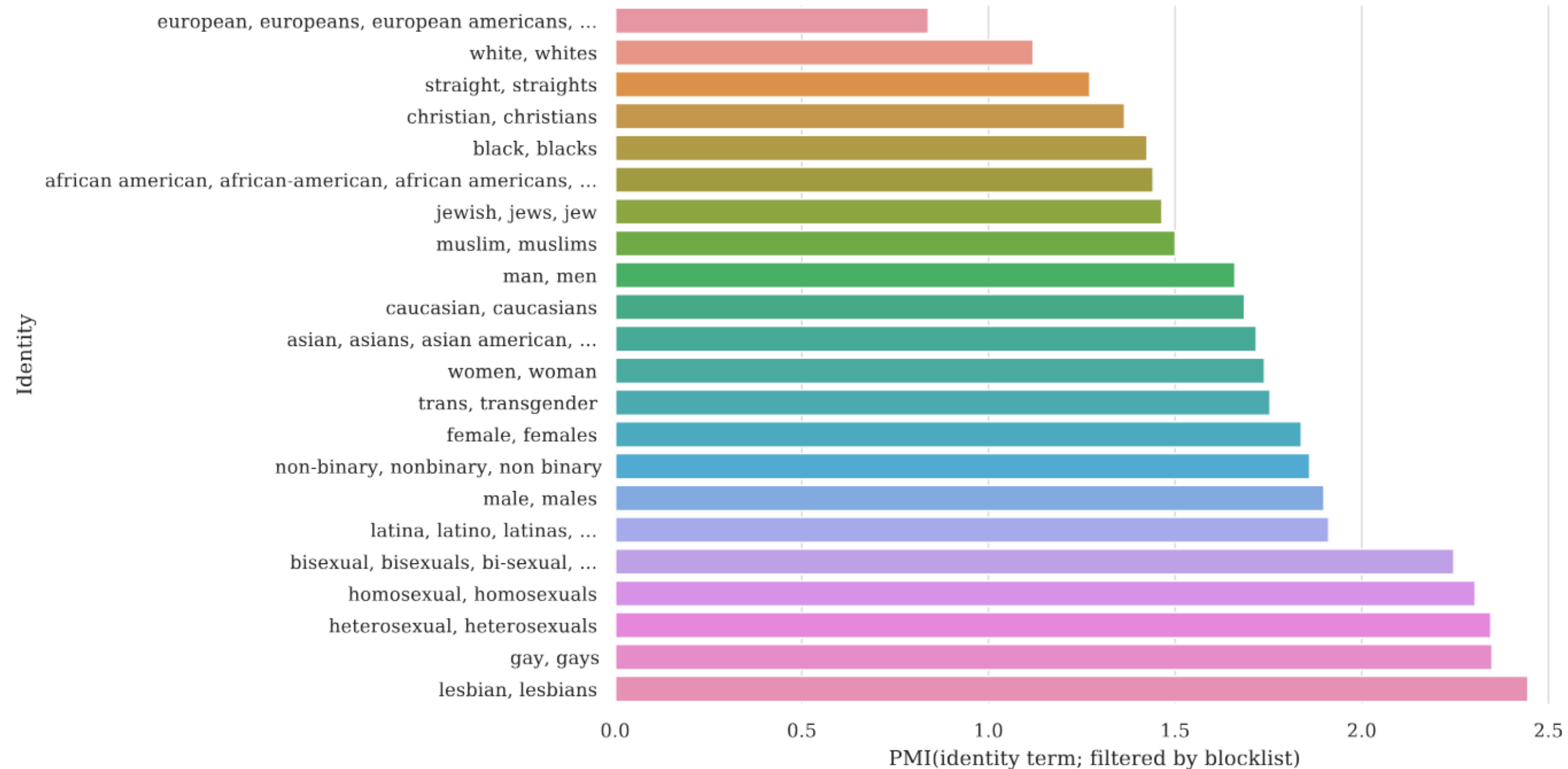


Figure 5: Pointwise Mutual Information (PMI) between identity mentions and documents being filtered out by the blacklist. Identities with higher PMI (e.g., lesbian, gay) have higher likelihood of being filtered out.

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



15 MINUTE READ



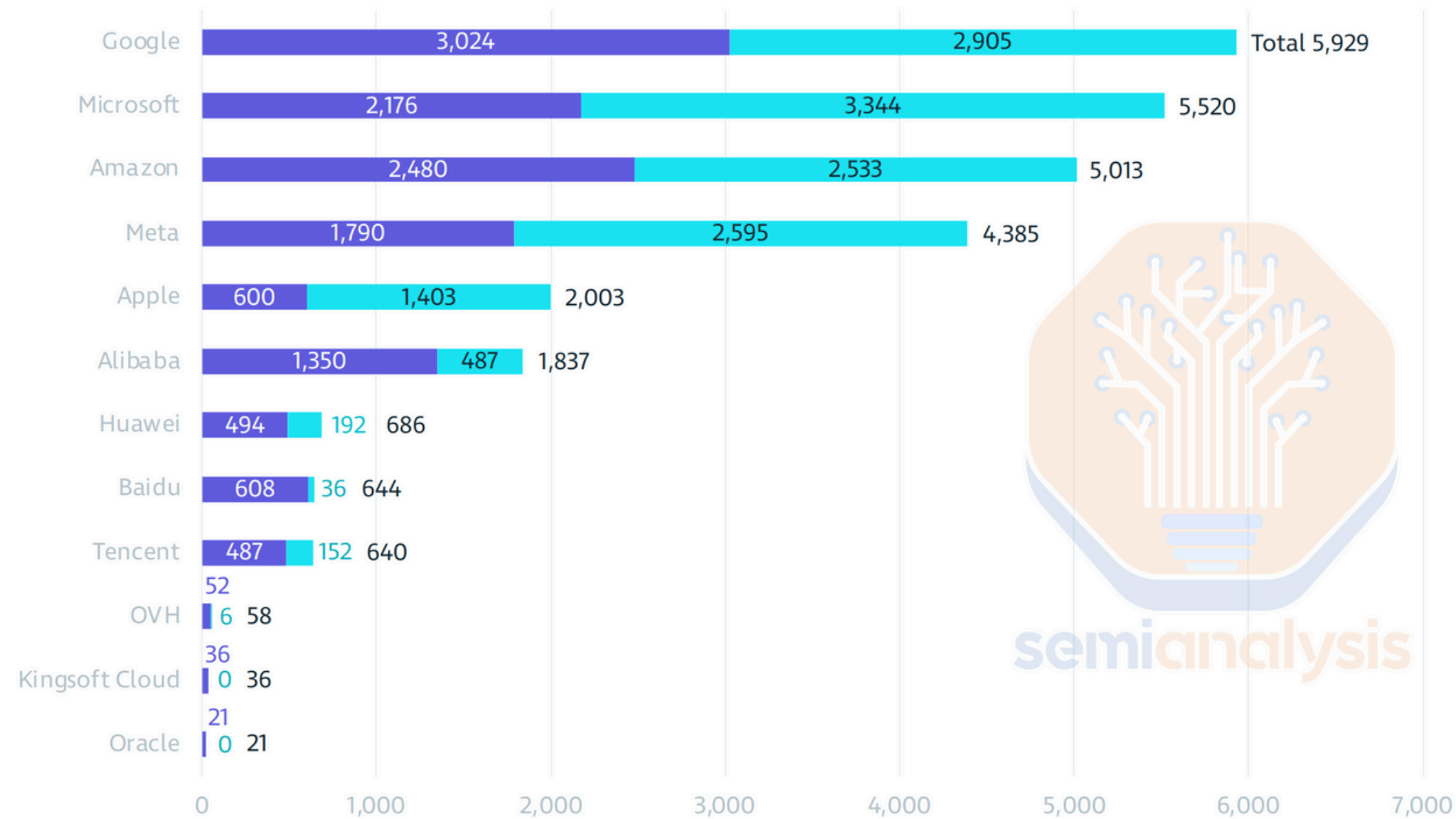
The map displays the global distribution of 15 companies, represented by black factory icons and blue dots. The companies are connected by lines, indicating relationships or data flow. The companies are: Telus Digital (Canada), Sama (USA), Hugo Inc. (USA), Cloud Factory (USA), StepWise (Mexico), Marjorel (Belgium), Teleperformance (Spain), Techno Brain Group (Saudi Arabia), iSON Experiences (Saudi Arabia), and several others in Africa and South America. The map includes labels for various countries like Canada, USA, Mexico, Colombia, Peru, Brazil, Argentina, UK, Belgium, Spain, Turkey, Iran, Saudi Arabia, India, Nigeria, Ethiopia, Kenya, and South Africa.

- Workers

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

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

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



RESOURCE INTENSIVE

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


 | **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 >](#)


[Earn Trust >](#)

[Expand Opportunity >](#)

[Advance Sustainability >](#)



GLARING ACCOUNTABILITY GAP


 | **Research** [Our research](#) [Programs & events](#) [Connect & learn](#) [About](#) [Register: Research Forum](#)


AI For Good Lab


[Watch video >](#)

[Overview](#) [The Prompt](#) [AI for Good book](#) [People](#) [Publications](#) [Downloads](#) [Projects](#) [News & features](#)

Initiatives


[Fundamental Rights >](#)

[Earn Trust >](#)

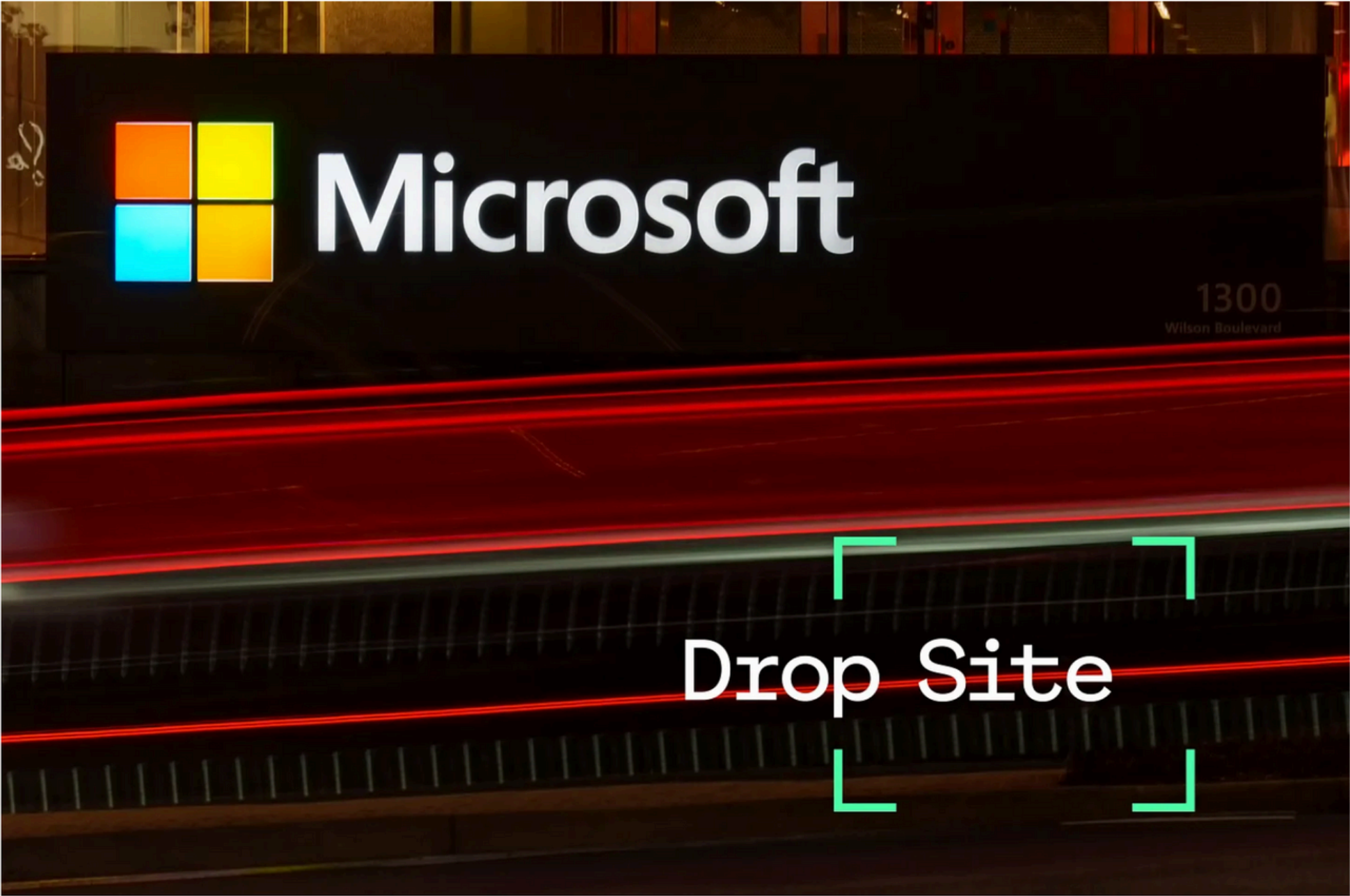
[Expand Opportunity >](#)

RYAN GRIM

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

 RYAN GRIM AND WAQAS AHMED
JAN 23, 2025

[459](#) [20](#) [102](#) [Share](#)



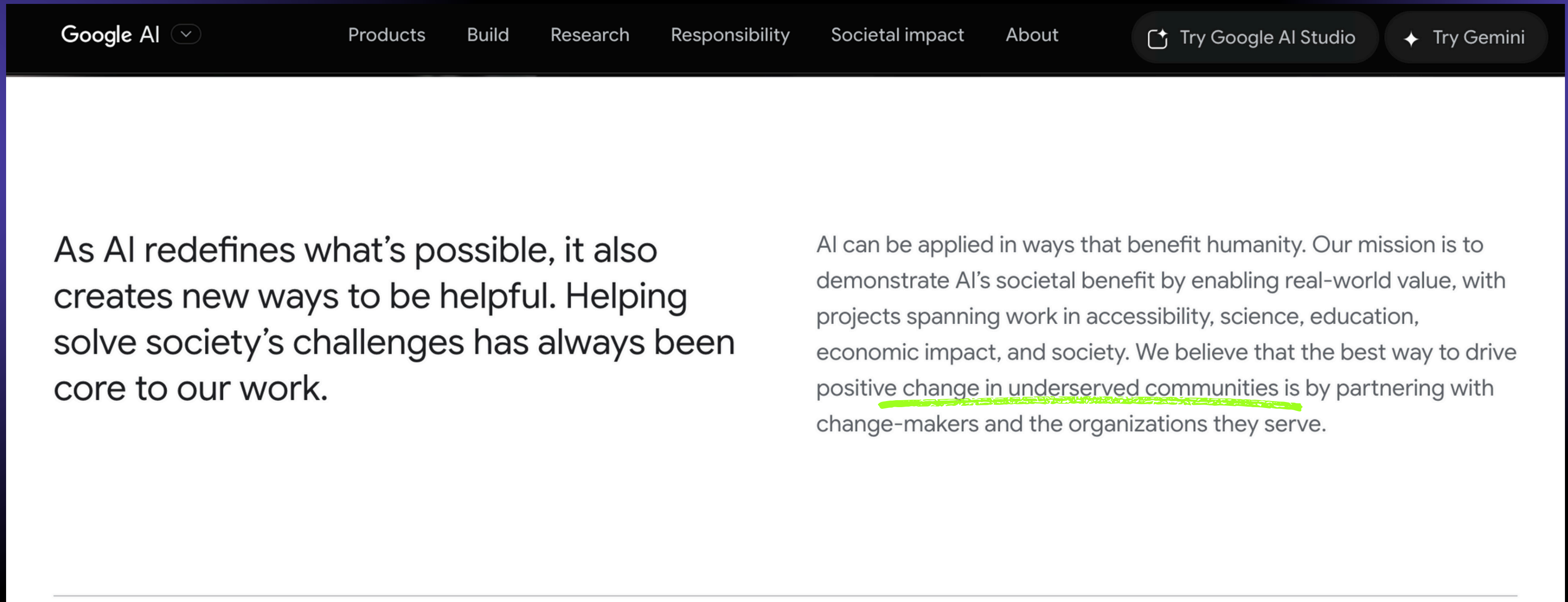
Drop Site

1300
Wilson Boulevard


Photo by J. David Ake/Getty Images



GLARING ACCOUNTABILITY GAP



The image is a screenshot of the Google AI website. The top navigation bar is dark with white text for 'Google AI' (with a dropdown arrow), 'Products', 'Build', 'Research', 'Responsibility', 'Societal impact', and 'About'. On the right side of the navigation bar are two buttons: 'Try Google AI Studio' and 'Try Gemini'. The main content area has a white background. On the left, there is a large text block. On the right, there is a smaller text block. The text on the right contains a highlighted phrase.

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 AI

Products

Build

Research

Responsibility

Societal impact

About

Try Google AI Studio

Try Gemini

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



Big tech and AI companies that enable war crimes, 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:

UNethical.

UNlawful.

UNacceptable.

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.



Microsoft



Palantir



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

