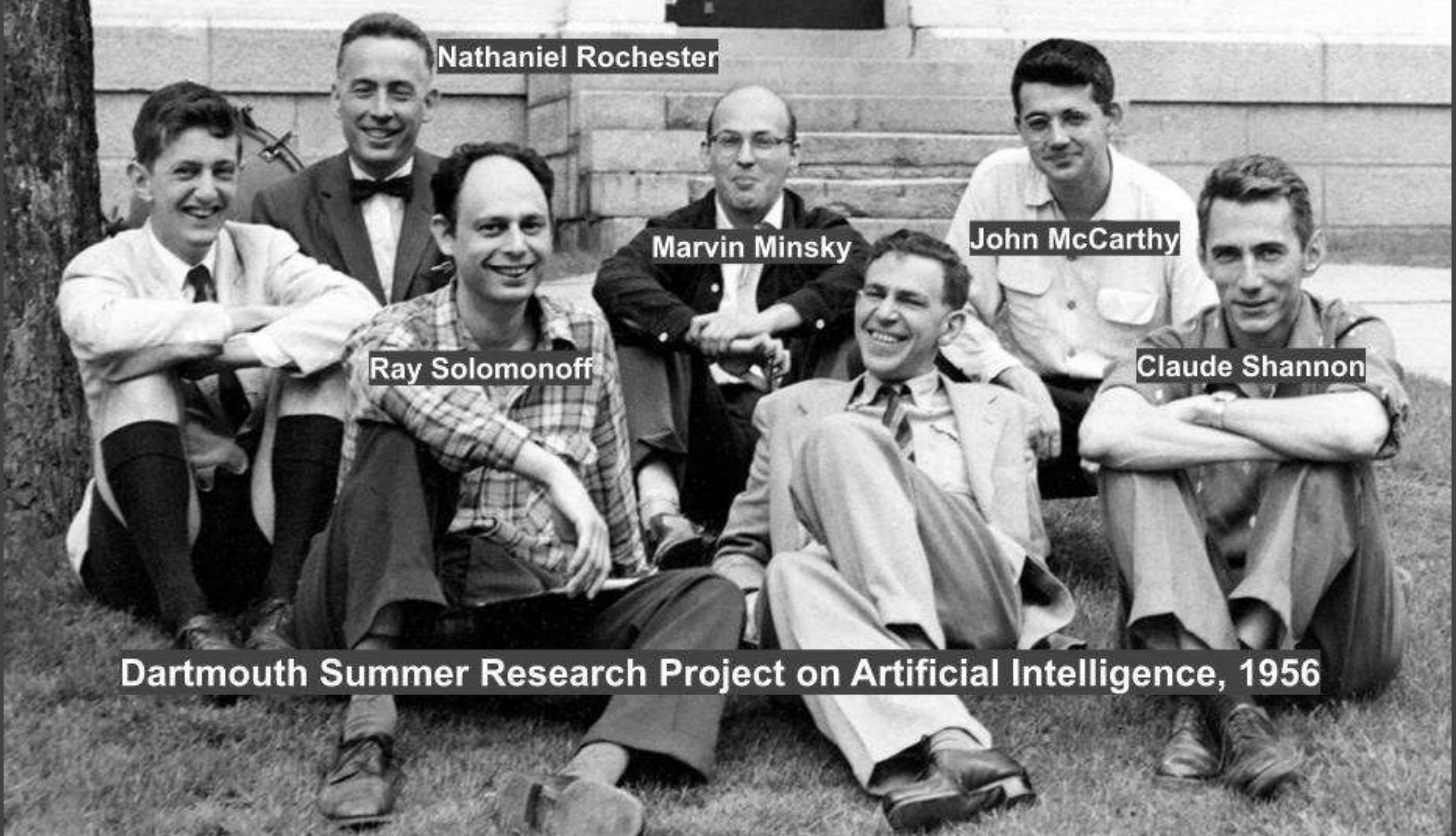
Etica e AI

Guido Boella

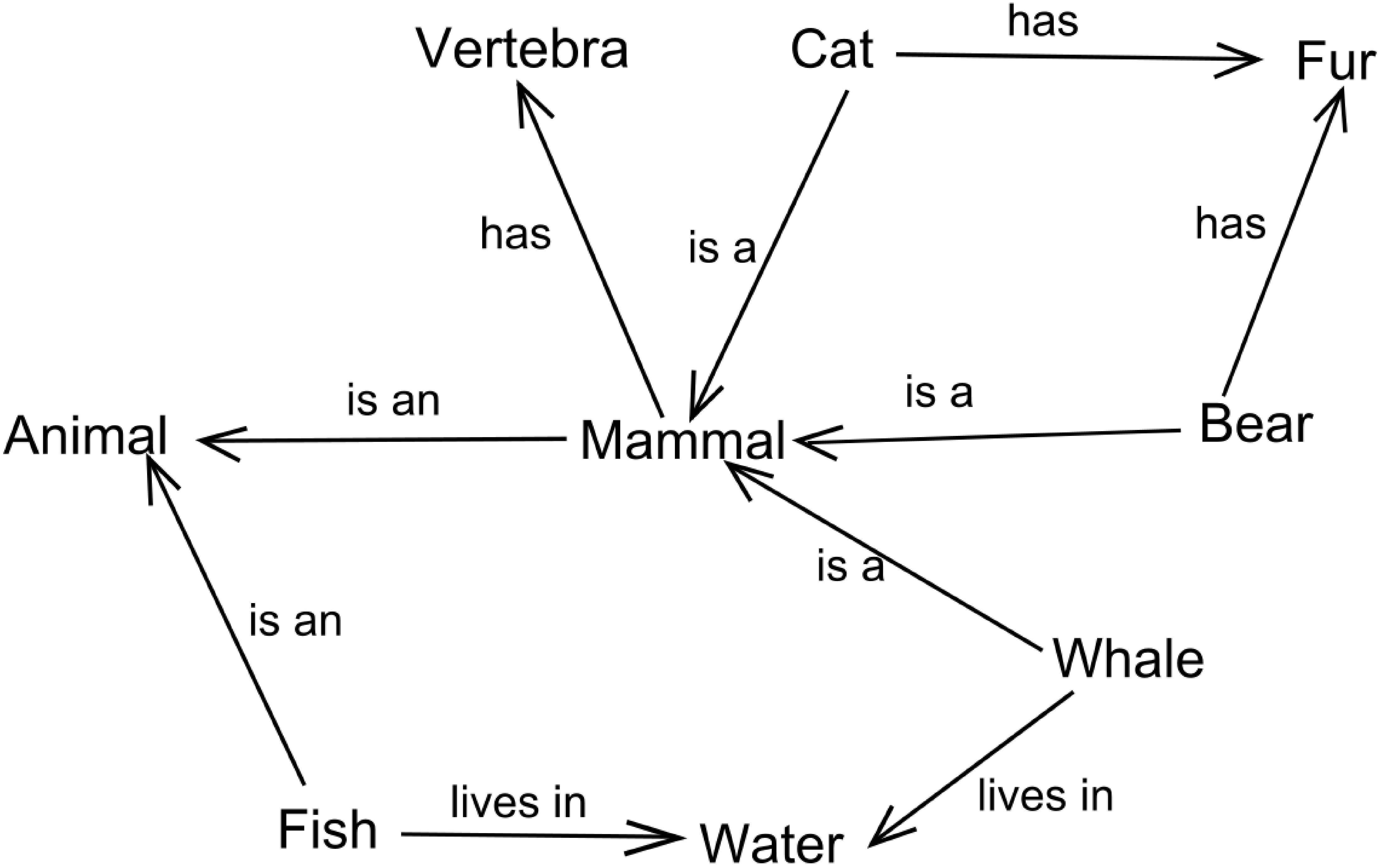
Dipartimento di Informatica – Università di Torino

Società Italiana per l'Etica dell'Intelligenza Artificiale
www.sipeia.it

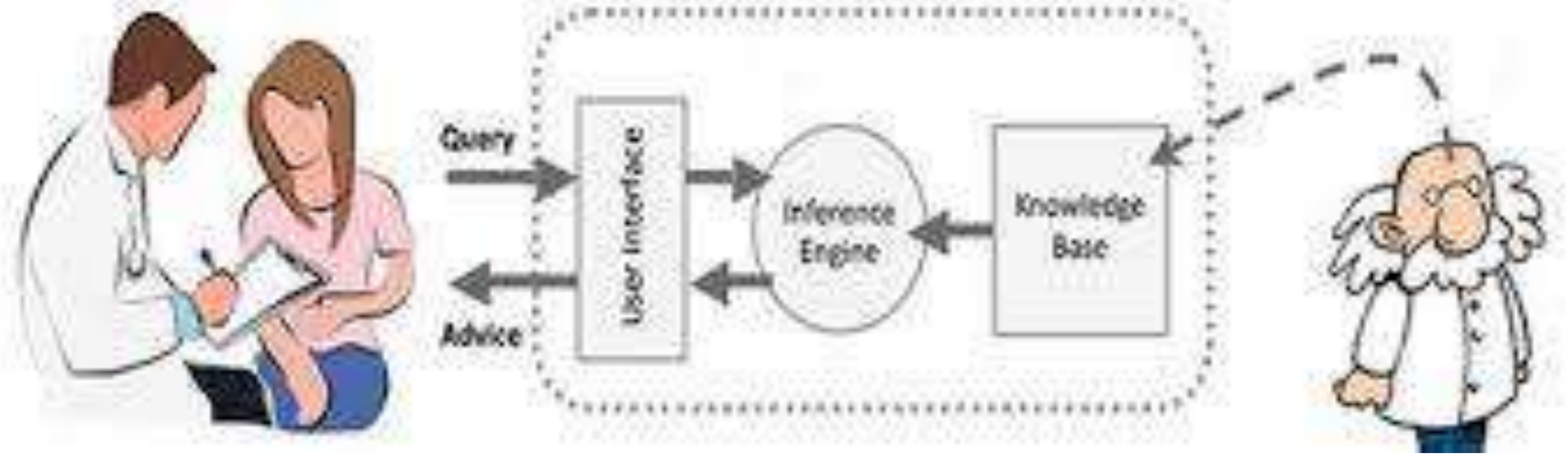
Progetto AI Aware www.ai-aware.eu

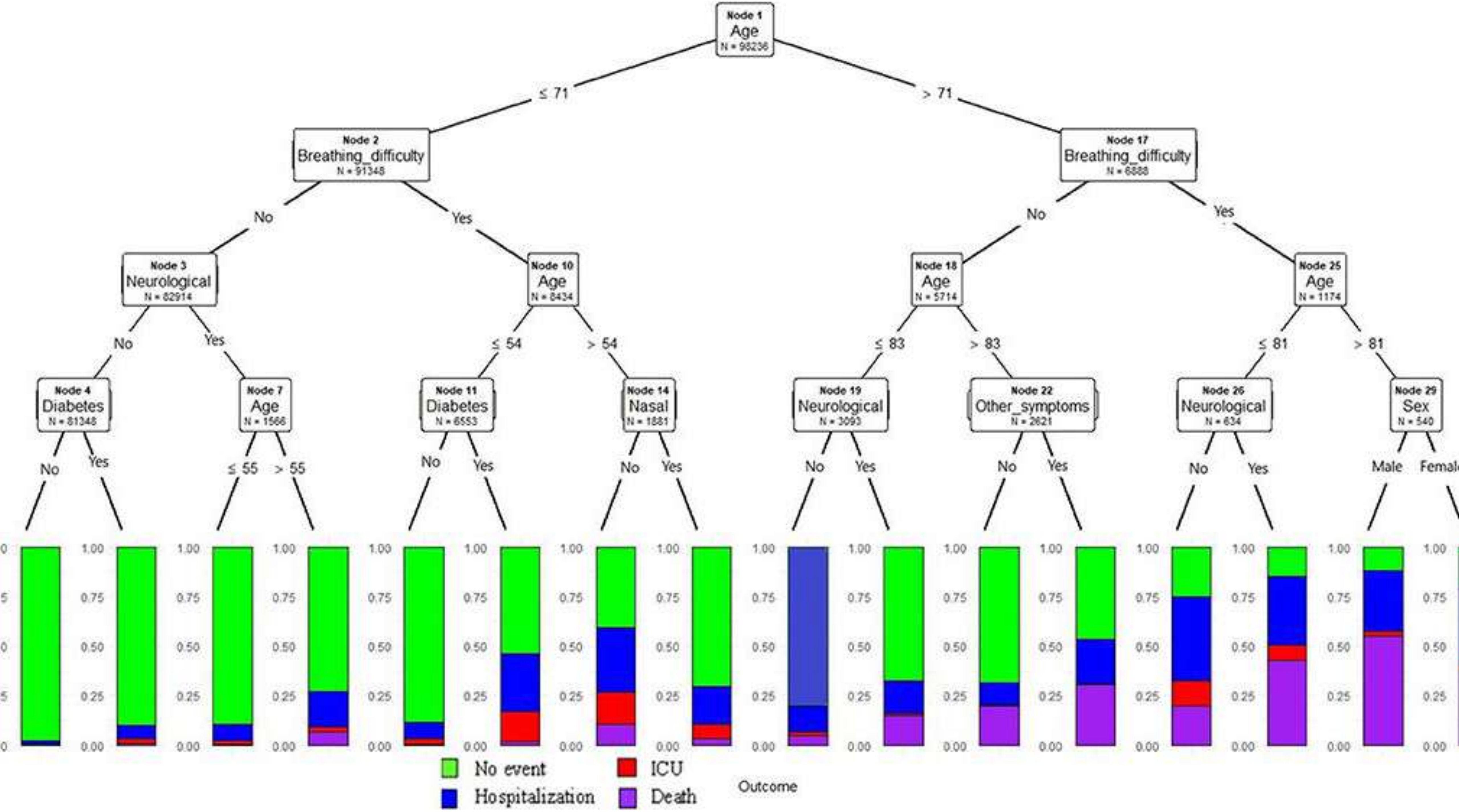


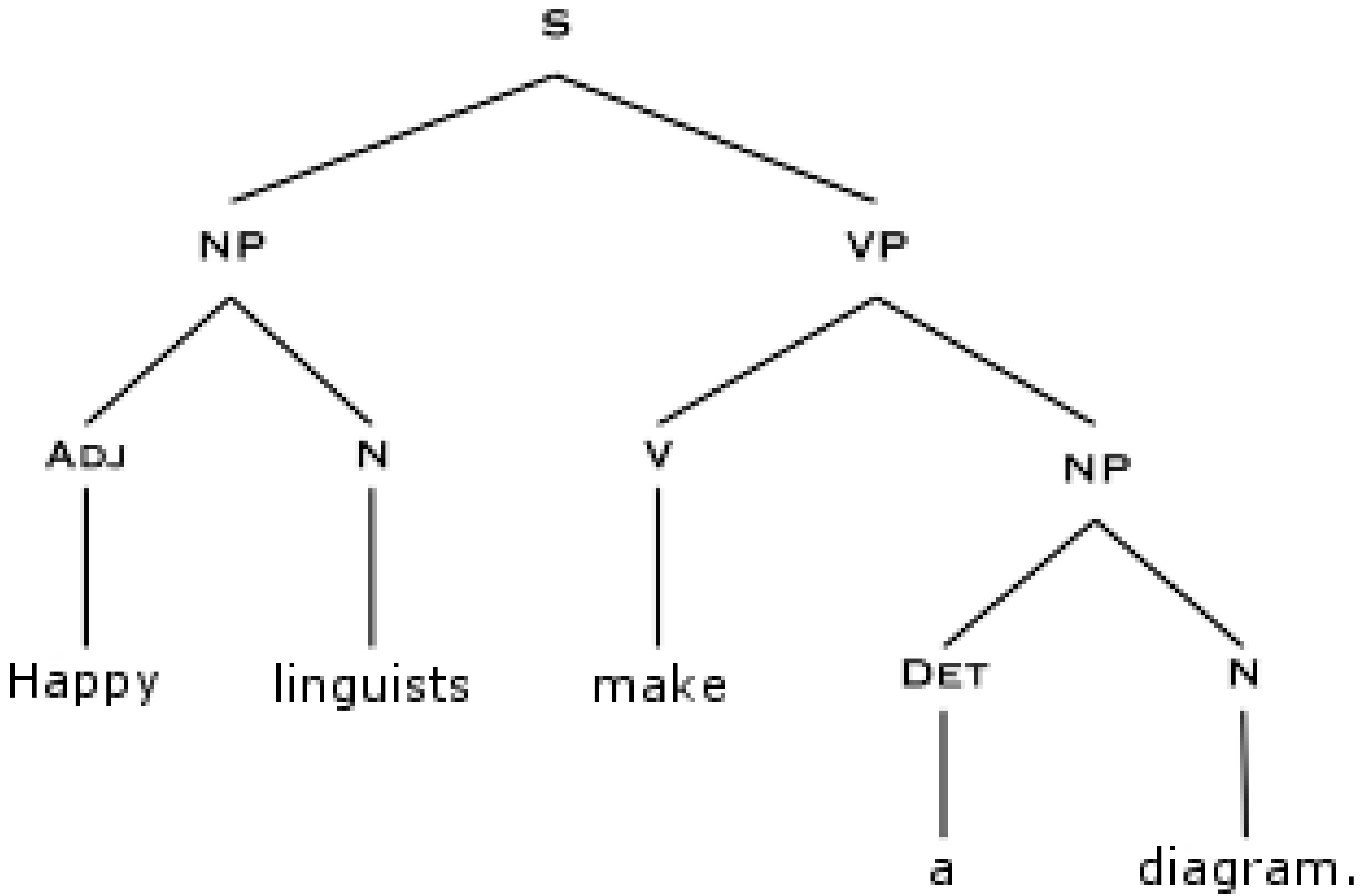
Dartmouth Summer Research Project on Artificial Intelligence, 1956

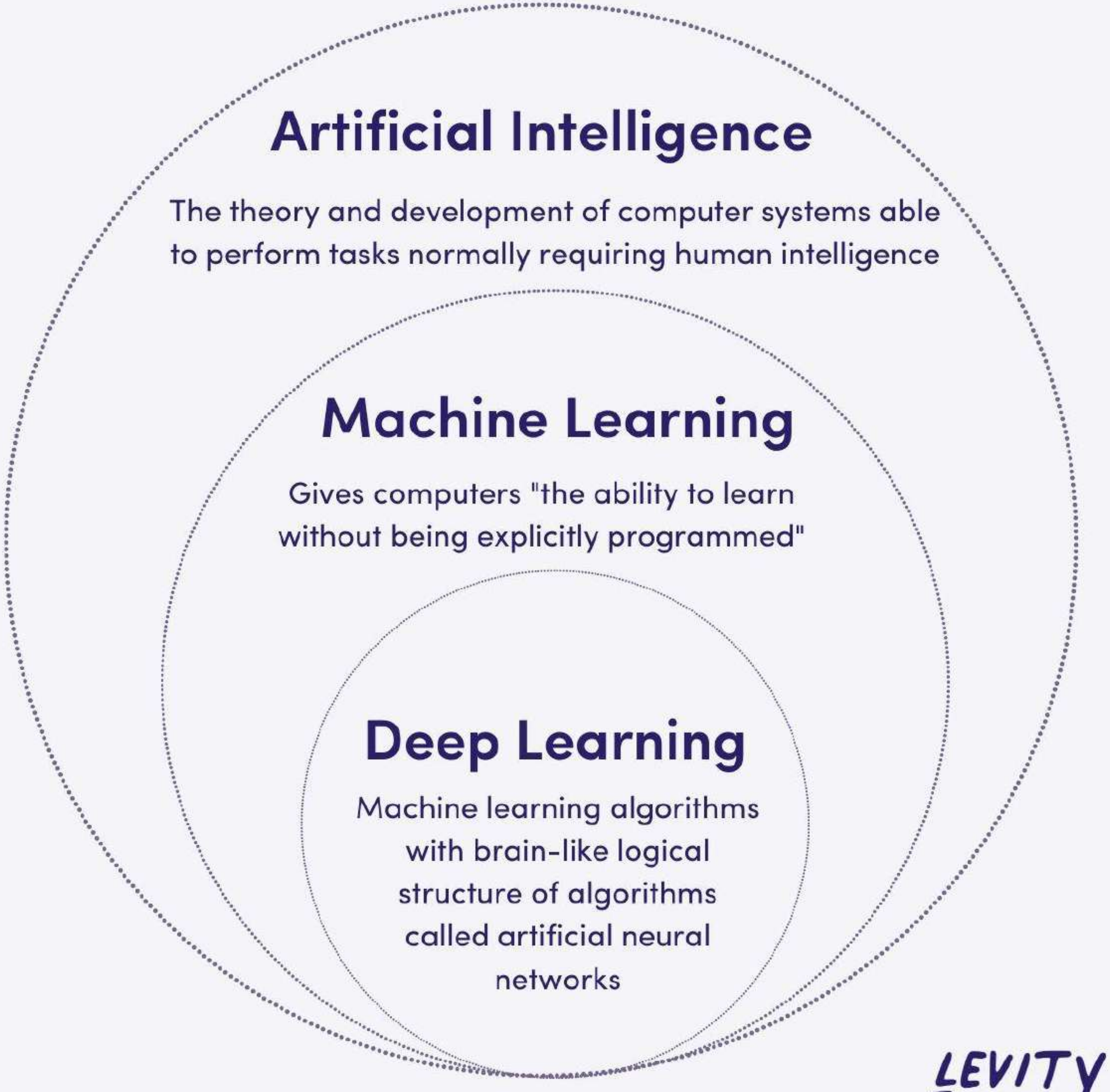

Vertebra
Cat
has
Fur
has
is a
has
is an
Animal
Mammal
is a
Bear
is a
is an
Whale
lives in
Fish
lives in
Water

Expert System




Node 1
Age
N = 98236
≤ 71
> 71
Node 2
Breathing_difficulty
N = 91348
No
Yes
Node 17
Breathing_difficulty
N = 6888
No
Yes
Node 3
Neurological
N = 82914
Node 10
Age
N = 8434
Node 18
Age
N = 5714
Node 25
Age
N = 1174
No
Yes
≤ 54
> 54
≤ 83
> 83
≤ 81
> 81
Node 4
Diabetes
N = 81348
Node 7
Age
N = 1566
Node 11
Diabetes
N = 6553
Node 14
Nasal
N = 1881
Node 19
Neurological
N = 3093
Node 22
Other_symptoms
N = 2621
Node 26
Neurological
N = 634
Node 29
Sex
N = 540
No
Yes
≤ 55
> 55
No
Yes
No
Yes
No
Yes
No
Yes
No
Yes
Male
Femal
1.00
0.75
0.50
0.25
0.00
No event
Hospitalization
ICU
Death
Outcome


S
NP
VP
Adj
N
V
NP
Happy
linguists
make
Det
N
a
diagram.



Artificial Intelligence

The theory and development of computer systems able to perform tasks normally requiring human intelligence

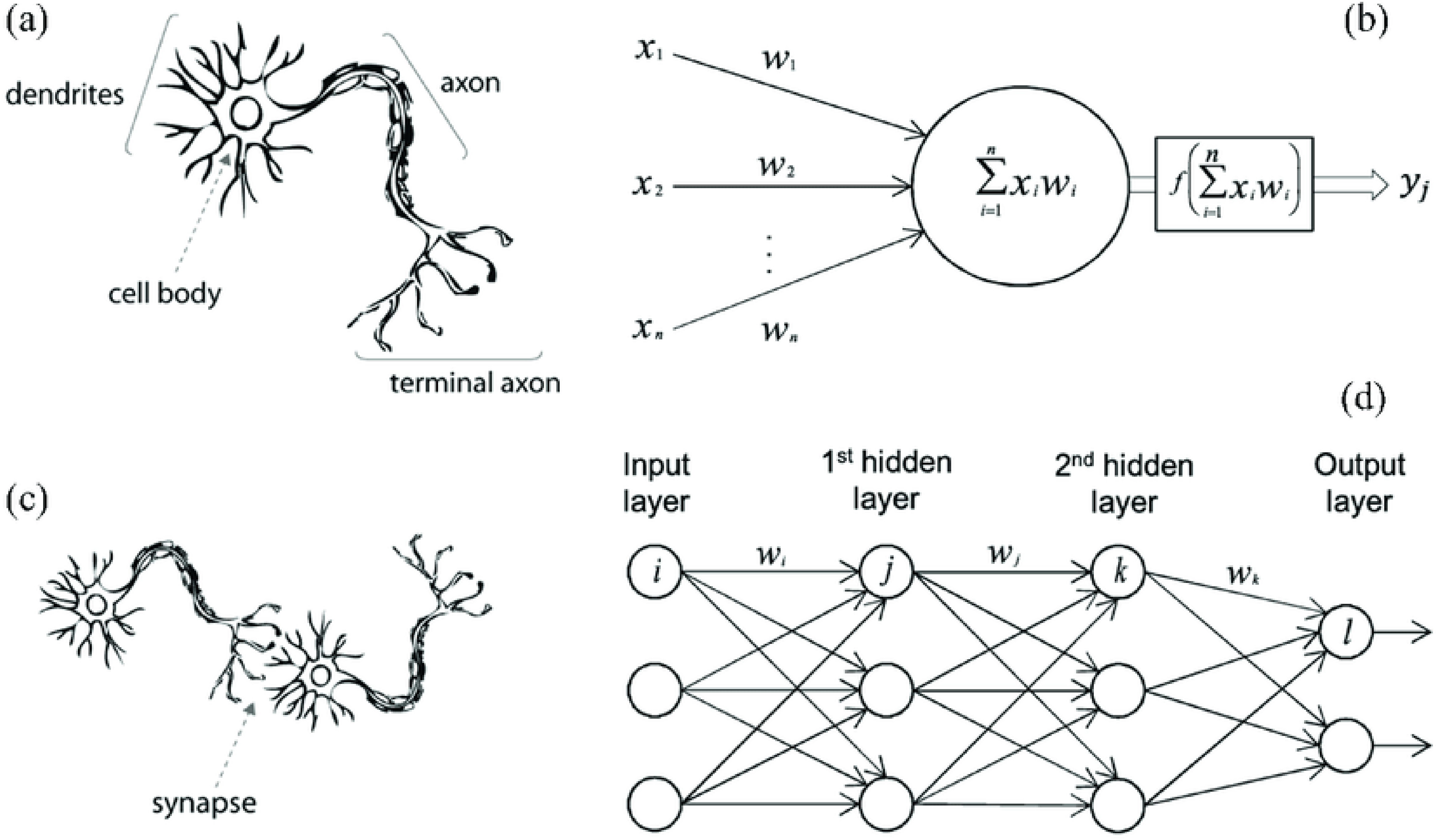
Machine Learning

Gives computers "the ability to learn without being explicitly programmed"

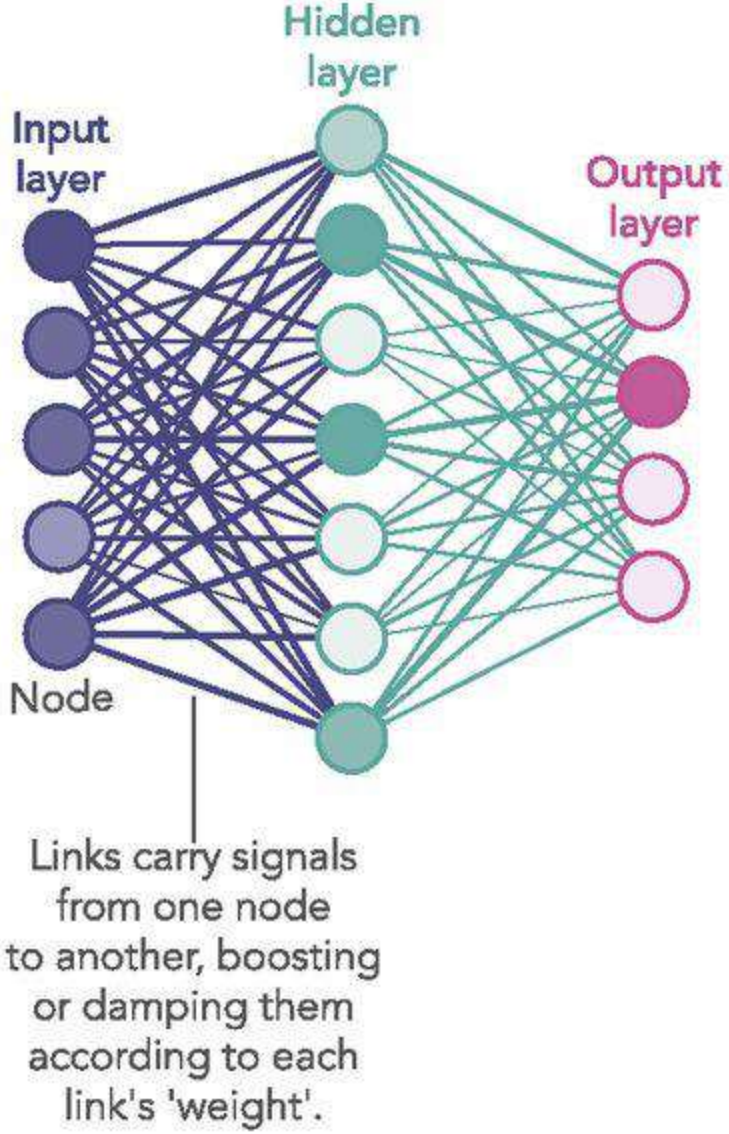
Deep Learning

Machine learning algorithms with brain-like logical structure of algorithms called artificial neural networks

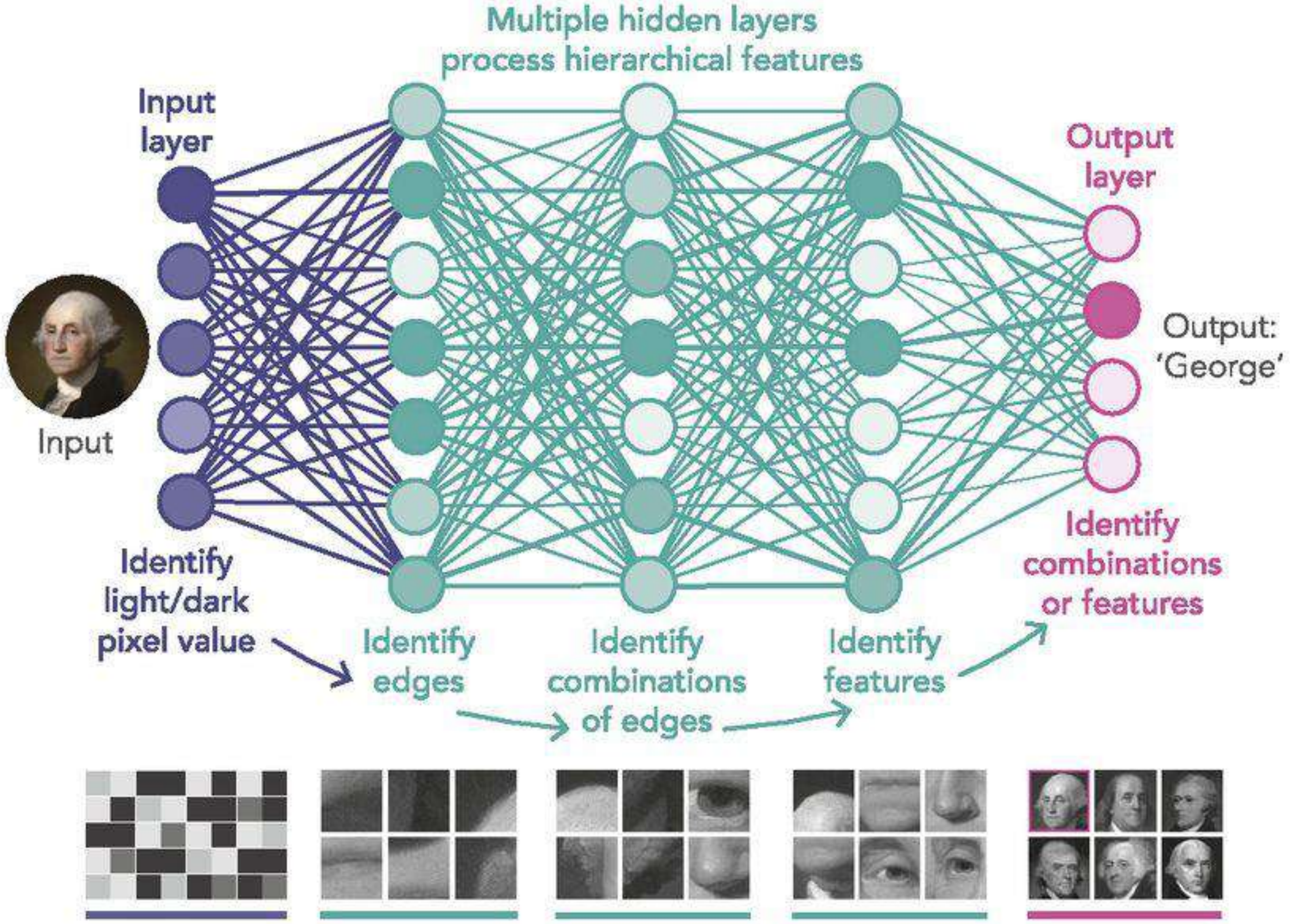
LEVITY

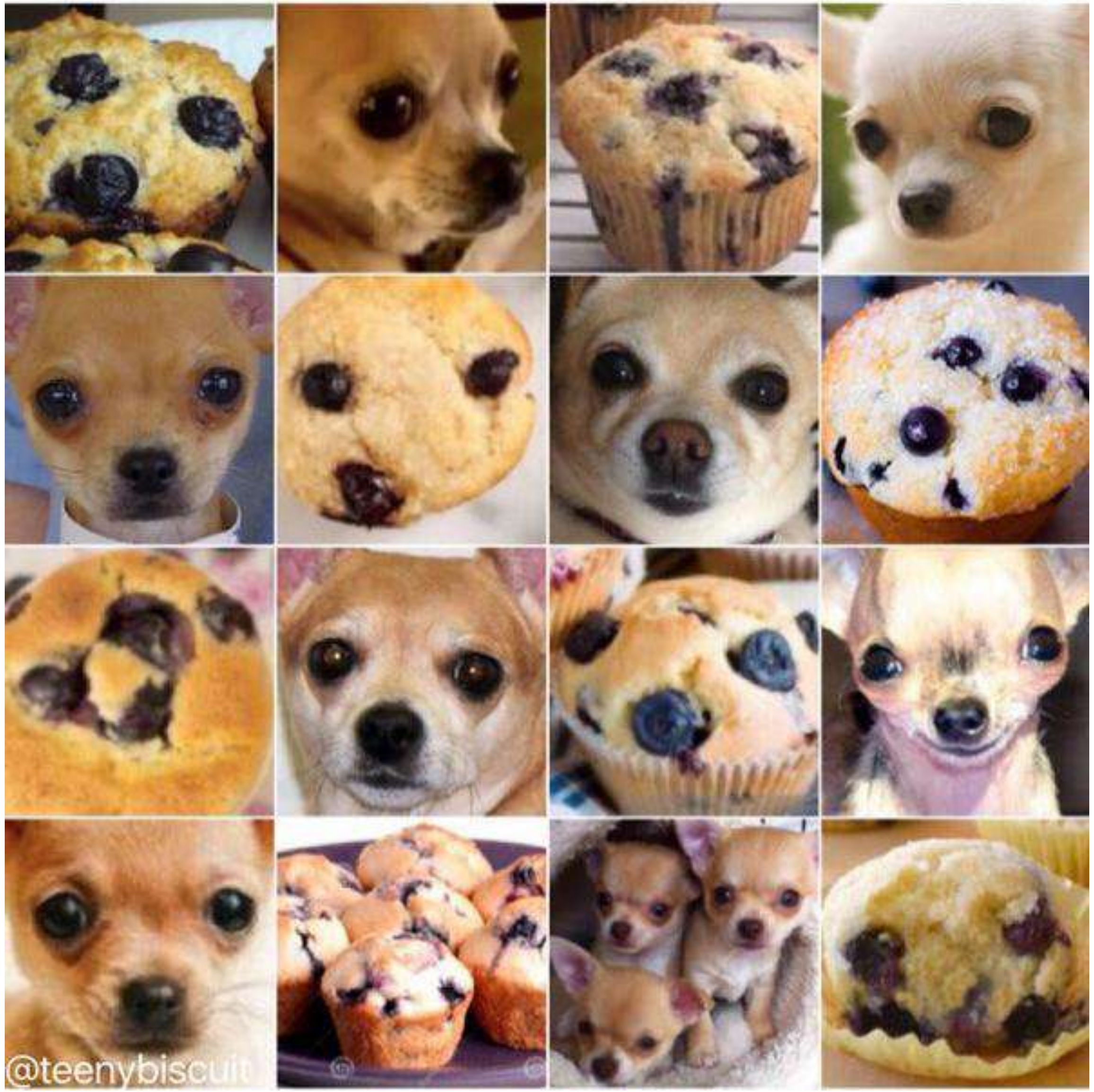

(a)
dendrites
axon
cell body
terminal axon
(b)
x_1
w_1
x_2
w_2
x_n
w_n
$\sum_{i=1}^{n} x_i w_i$
$f\left(\sum_{i=1}^{n} x_i w_i\right)$
y_j
(c)
synapse
(d)
Input layer
1st hidden layer
2nd hidden layer
Output layer
i
w_i
j
w_j
k
w_k
l

1980S-ERA NEURAL NETWORK


Hidden layer
Input layer
Output layer
Node
Links carry signals from one node to another, boosting or damping them according to each link's 'weight'.

DEEP LEARNING NEURAL NETWORK


Multiple hidden layers process hierarchical features
Input layer
Output layer
Input
Output: 'George'
Identify light/dark pixel value
Identify edges
Identify combinations of edges
Identify features
Identify combinations or features


@teenybiscuit

How AI systems amplify bias

Image recognition systems that use biased machine learning data sets will inadvertently magnify that bias. Researchers are examining ways to reduce the effects.

COOKING

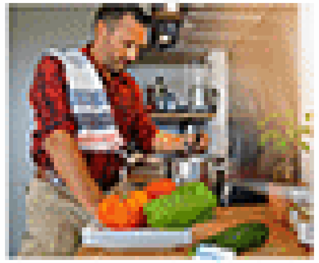
ROLE		VALUE
AGENT	►	WOMAN
FOOD	►	PASTA
HEAT	►	STOVE
TOOL	►	SPATULA
PLACE	►	KITCHEN

COOKING

ROLE		VALUE
AGENT	►	WOMAN
FOOD	►	FRUIT
HEAT	►	–
TOOL	►	KNIFE
PLACE	►	KITCHEN

COOKING

ROLE		VALUE
AGENT	►	WOMAN
FOOD	►	MEAT
HEAT	►	GRILL
TOOL	►	TONGS
PLACE	►	OUTSIDE



COOKING

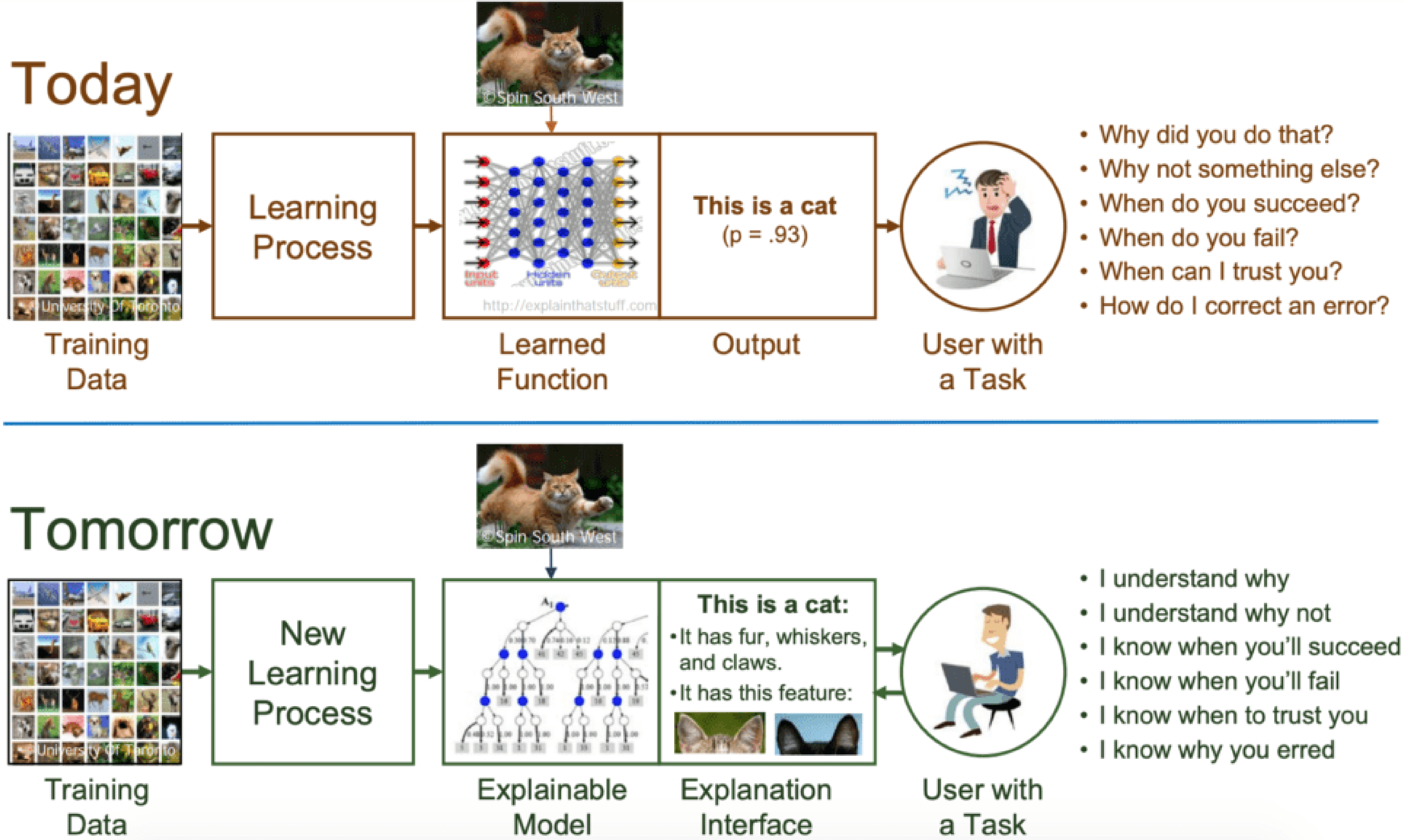
ROLE		VALUE
AGENT	►	WOMAN
FOOD	►	VEGETABLES
HEAT	►	STOVE
TOOL	►	TONGS
PLACE	►	KITCHEN

COOKING

ROLE		VALUE
AGENT	►	MAN
FOOD	►	–
HEAT	►	STOVE
TOOL	►	SPATULA
PLACE	►	KITCHEN

In this example of gender bias, adapted from a report published by researchers from the University of Virginia and the University of Washington, a visual semantic role labeling system has learned to identify a person cooking as female, even when the image is male.

IMAGES: (FROM LEFT) ALEXEY STIOP/ADOBE STOCK; JACOB LUND/ADOBE STOCK; BERND JÜRGENS/ADOBE STOCK; GRANGER WOOTZ/GETTY IMAGES



Today

Training Data → Learning Process → Learned Function → Output → User with a Task

This is a cat (p = .93)

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Tomorrow

Training Data → New Learning Process → Explainable Model → Explanation Interface ⇄ User with a Task

This is a cat:

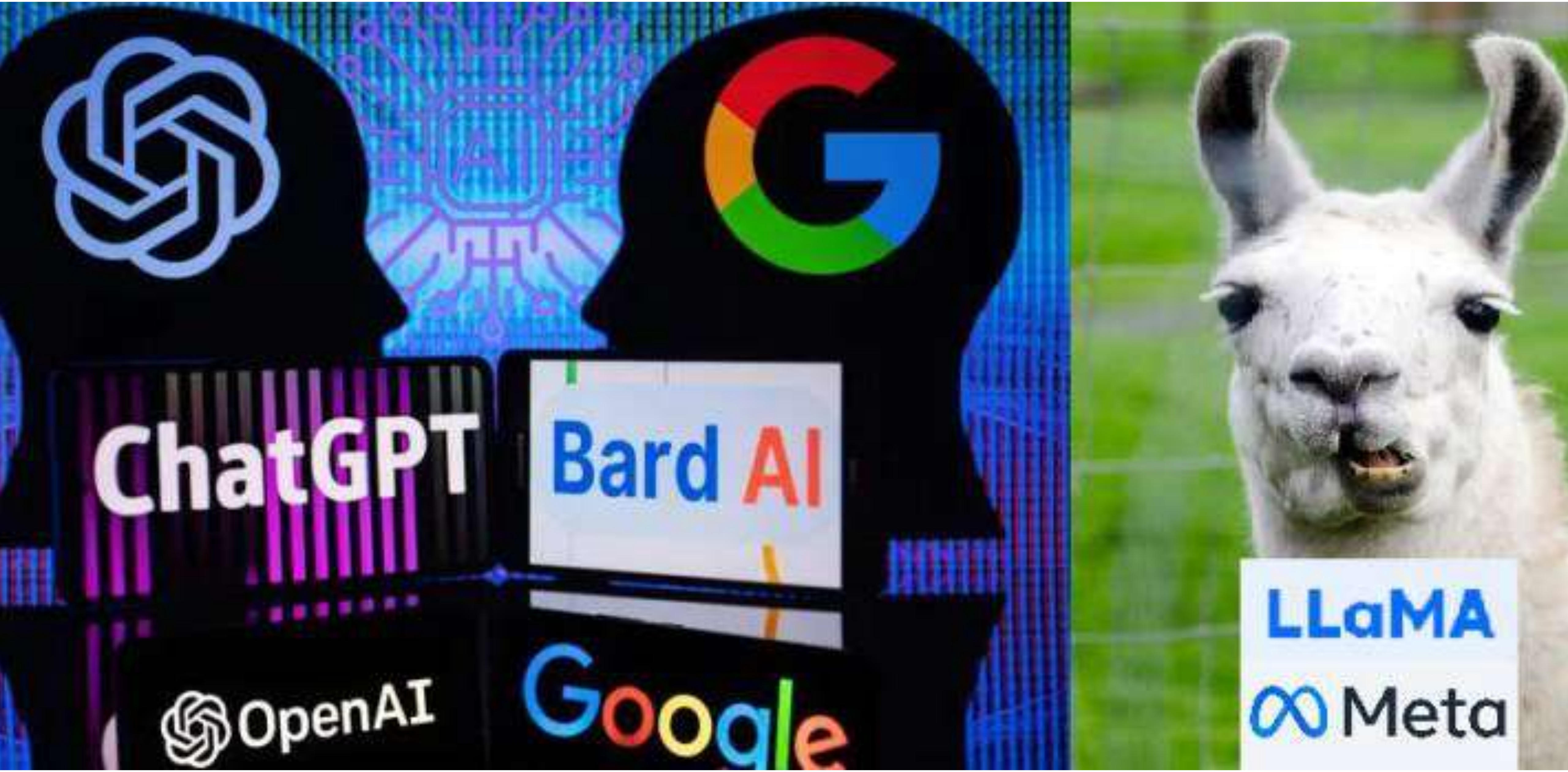
- It has fur, whiskers, and claws.
- It has this feature:

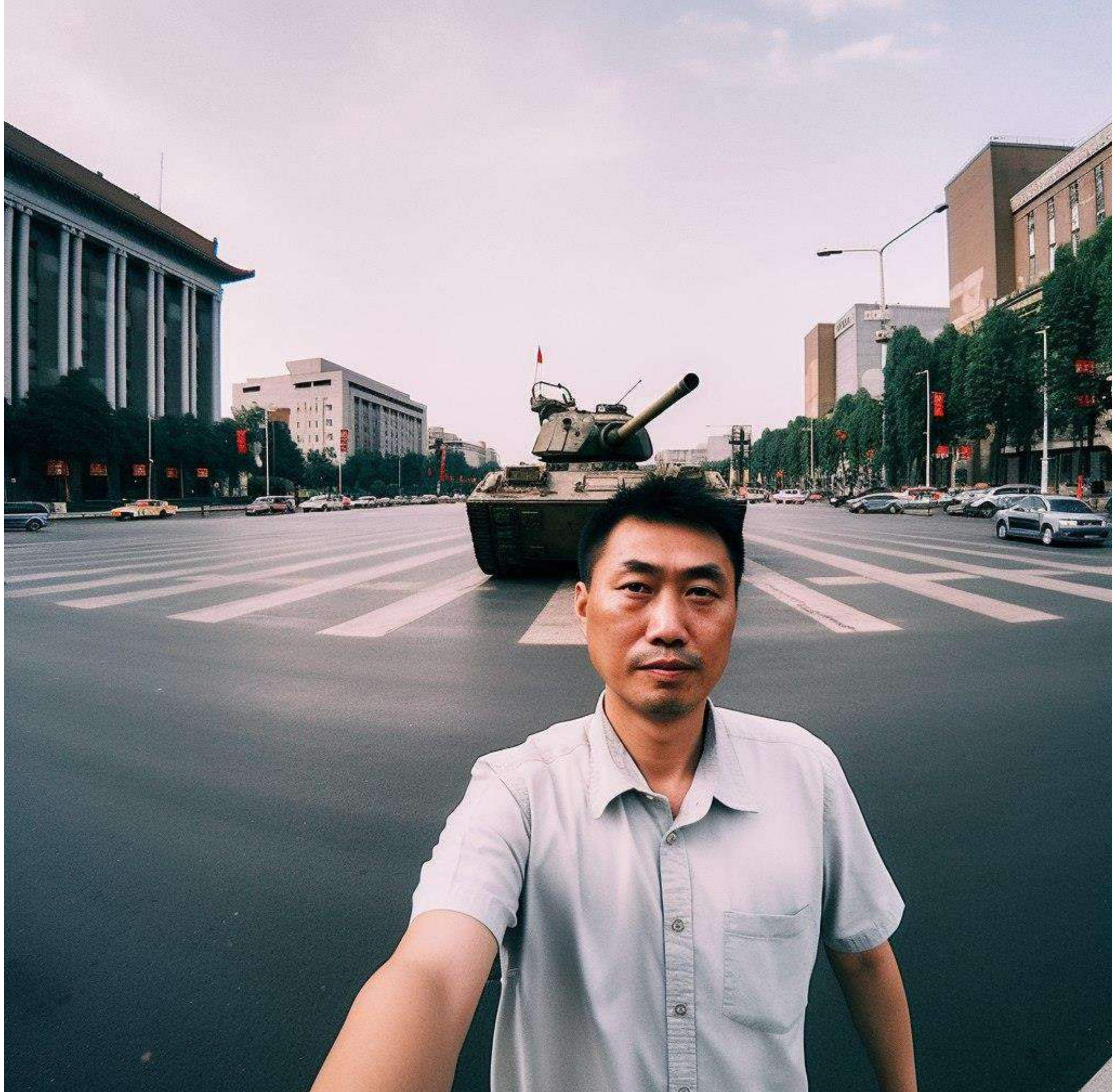
- I understand why
- I understand why not
- I know when you'll succeed
- I know when you'll fail
- I know when to trust you
- I know why you erred

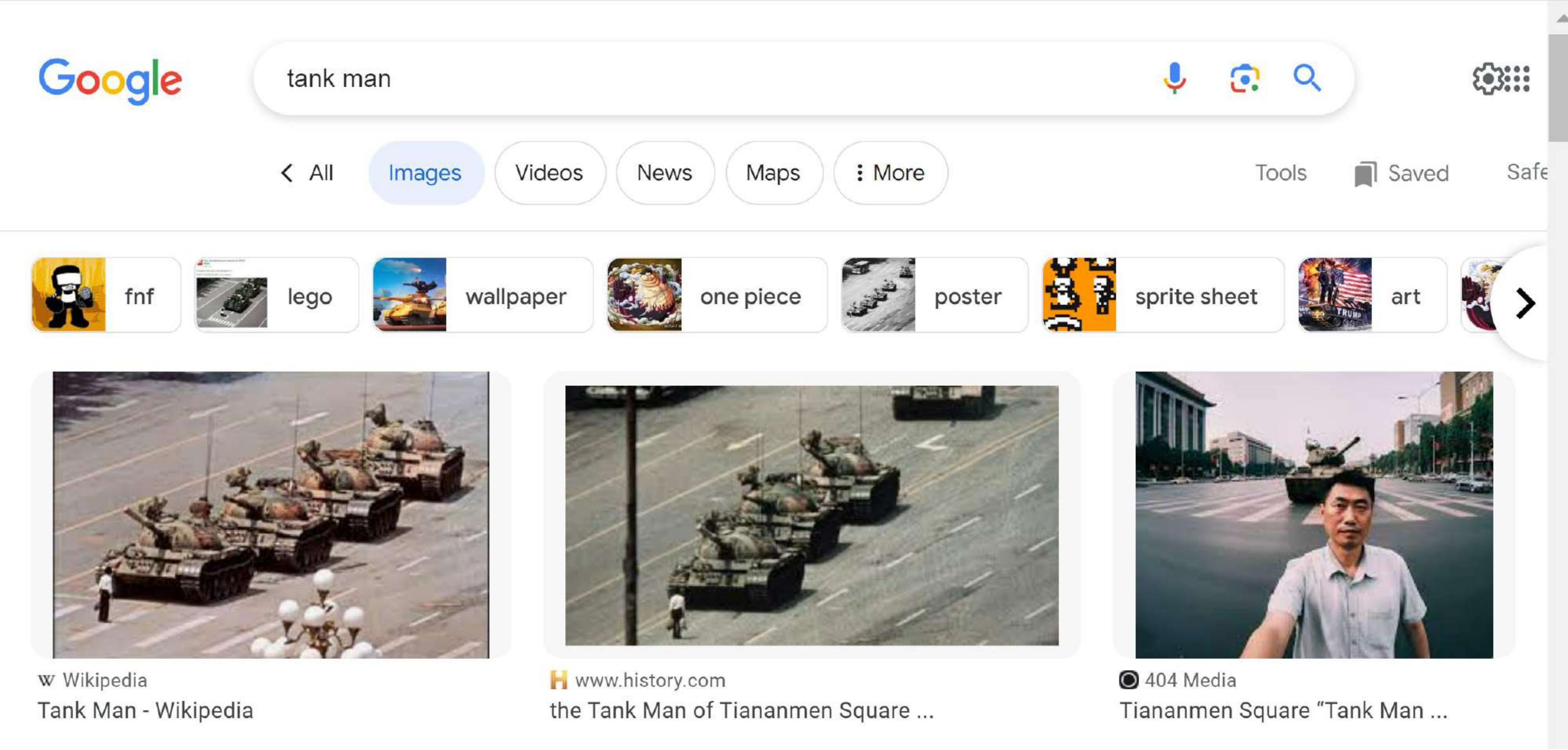
柯洁 KE JIE
00:46:57

LLM

LARGE LANGUAGE MODEL


AI
ChatGPT
Bard AI
OpenAI
Google
LLaMA
Meta



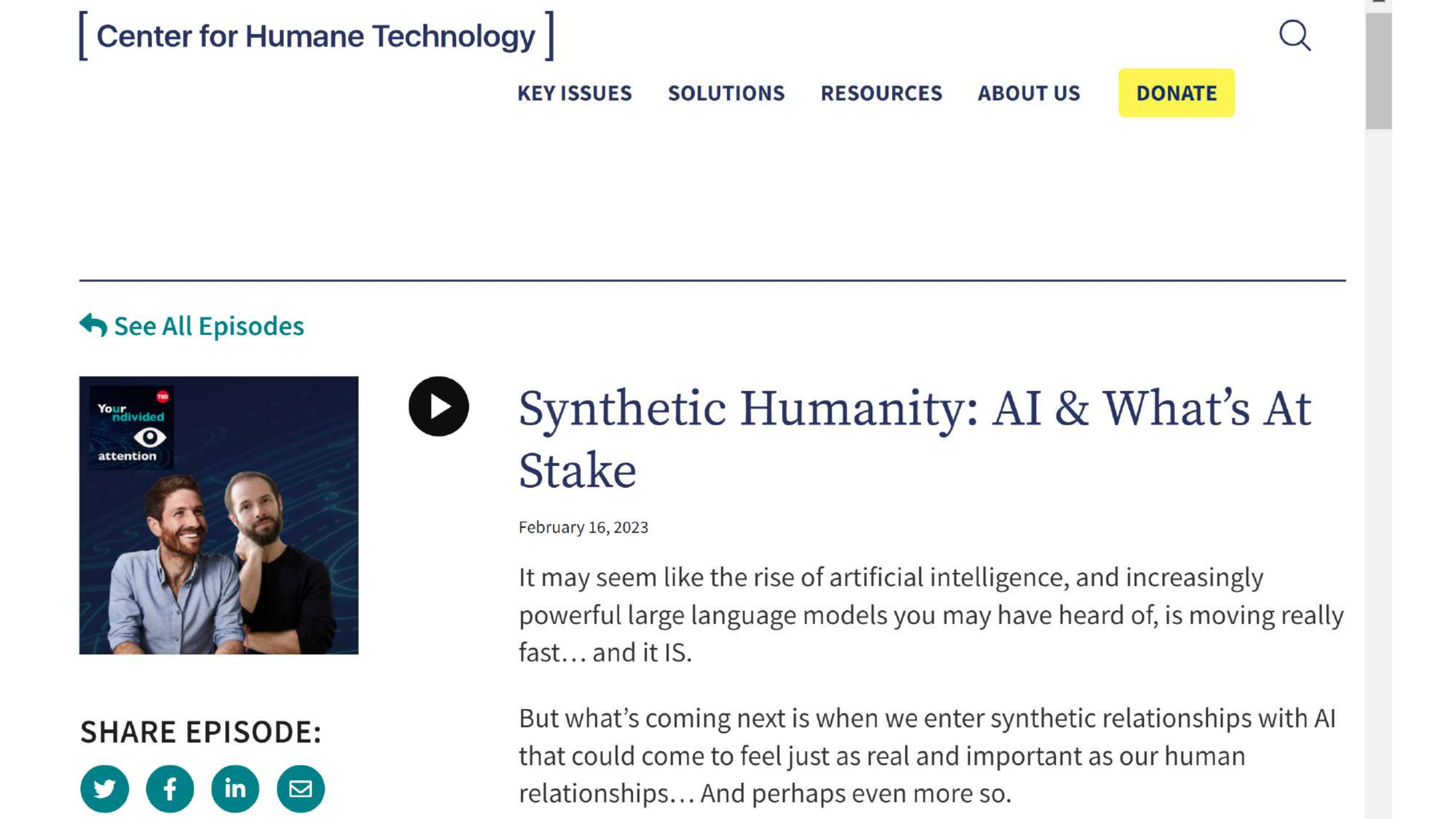


Wikipedia
Tank Man - Wikipedia

www.history.com
the Tank Man of Tiananmen Square ...

404 Media
Tiananmen Square “Tank Man ...

STABLE DIFFUSION
DALLE 2
MIDJOURNEY



See All Episodes

Synthetic Humanity: AI & What’s At Stake

February 16, 2023

It may seem like the rise of artificial intelligence, and increasingly powerful large language models you may have heard of, is moving really fast… and it IS.

But what’s coming next is when we enter synthetic relationships with AI that could come to feel just as real and important as our human relationships… And perhaps even more so.

SHARE EPISODE:

TECHNOLOGY

The Problem With Counterfeit People

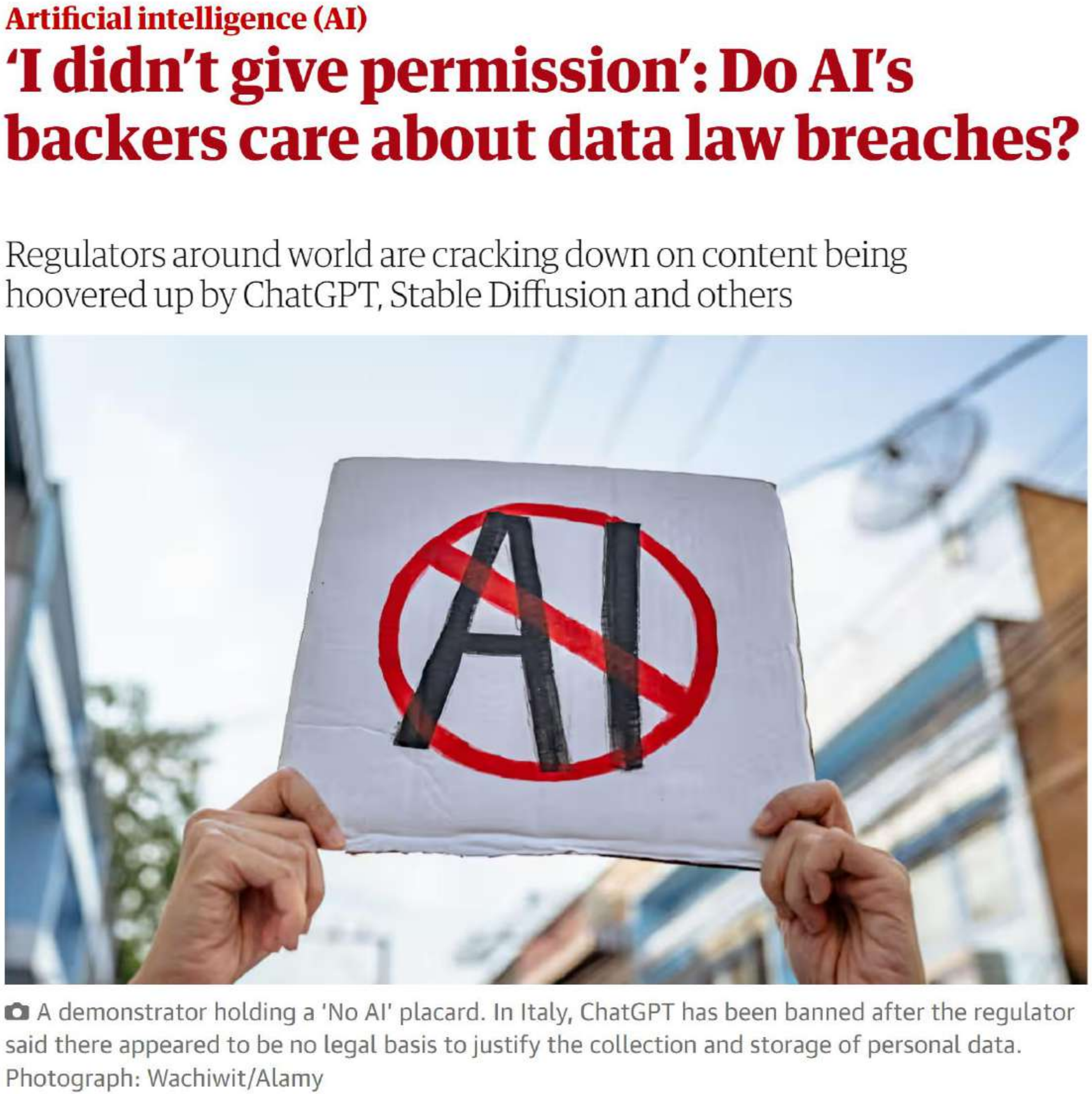
Companies using AI to generate fake people are committing an immoral act of vandalism, and should be held liable.

By Daniel C. Dennett

This Disinformation Is Just for You

Generative AI won't just flood the internet with more lies—it may also create convincing disinformation that's targeted at groups or even individuals.

PHOTOGRAPH: ROBERT BROOK/GETTY IMAGES

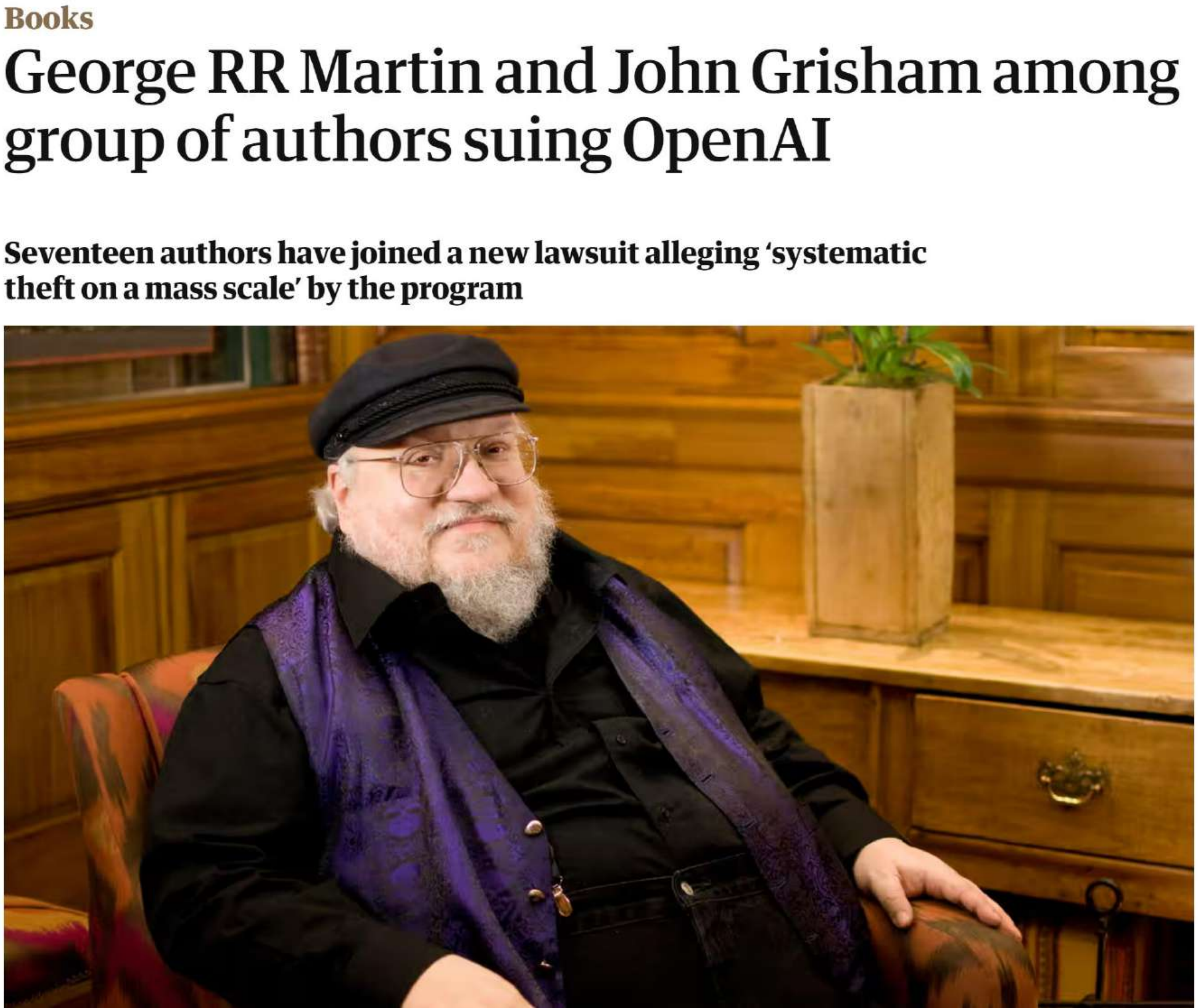


Artificial intelligence (AI)

‘I didn’t give permission’: Do AI’s backers care about data law breaches?

Regulators around world are cracking down on content being hoovered up by ChatGPT, Stable Diffusion and others

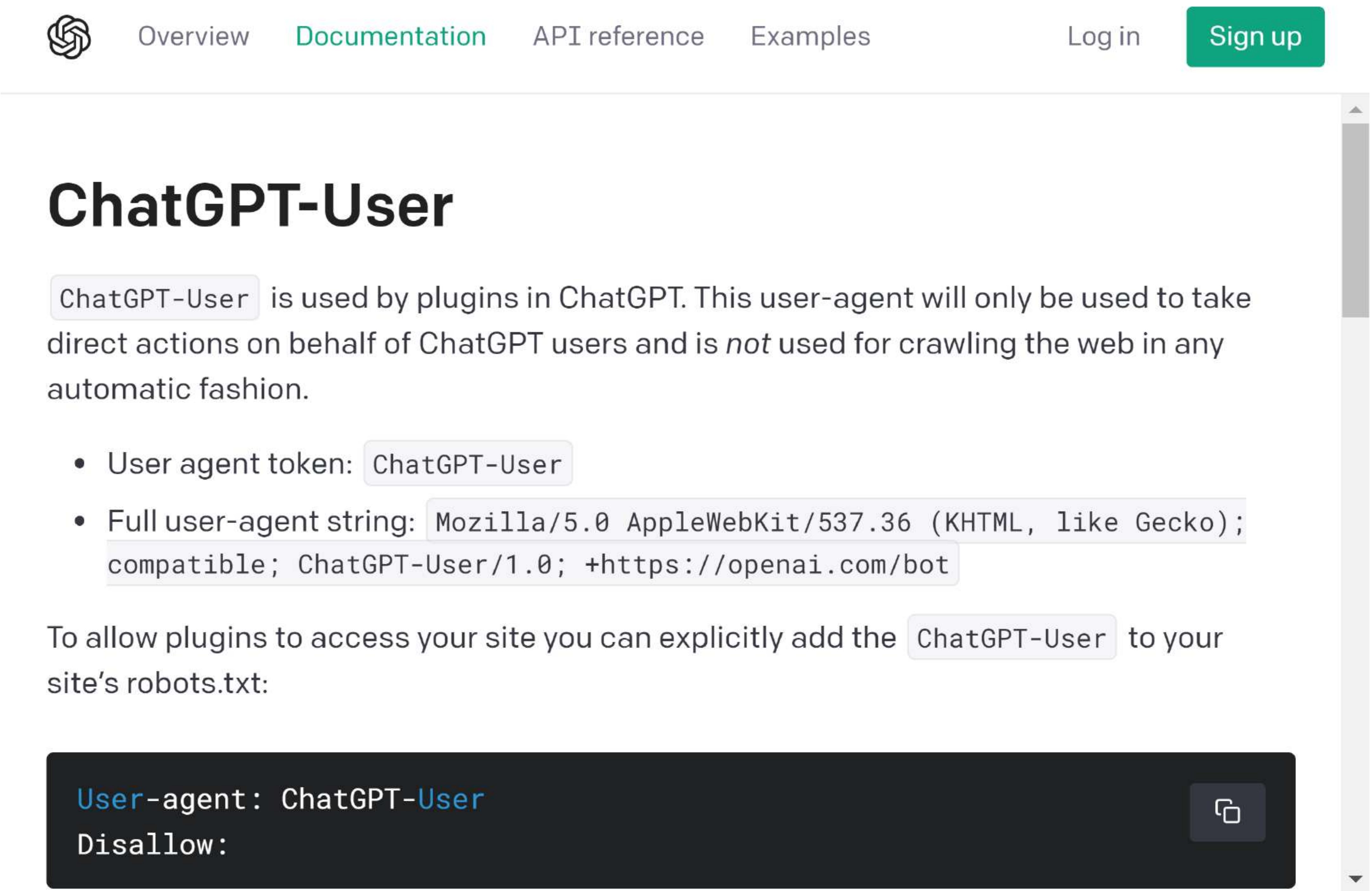
A demonstrator holding a ‘No AI’ placard. In Italy, ChatGPT has been banned after the regulator said there appeared to be no legal basis to justify the collection and storage of personal data. Photograph: Wachiwit/Alamy



Books

George RR Martin and John Grisham among group of authors suing OpenAI

Seventeen authors have joined a new lawsuit alleging 'systematic theft on a mass scale' by the program



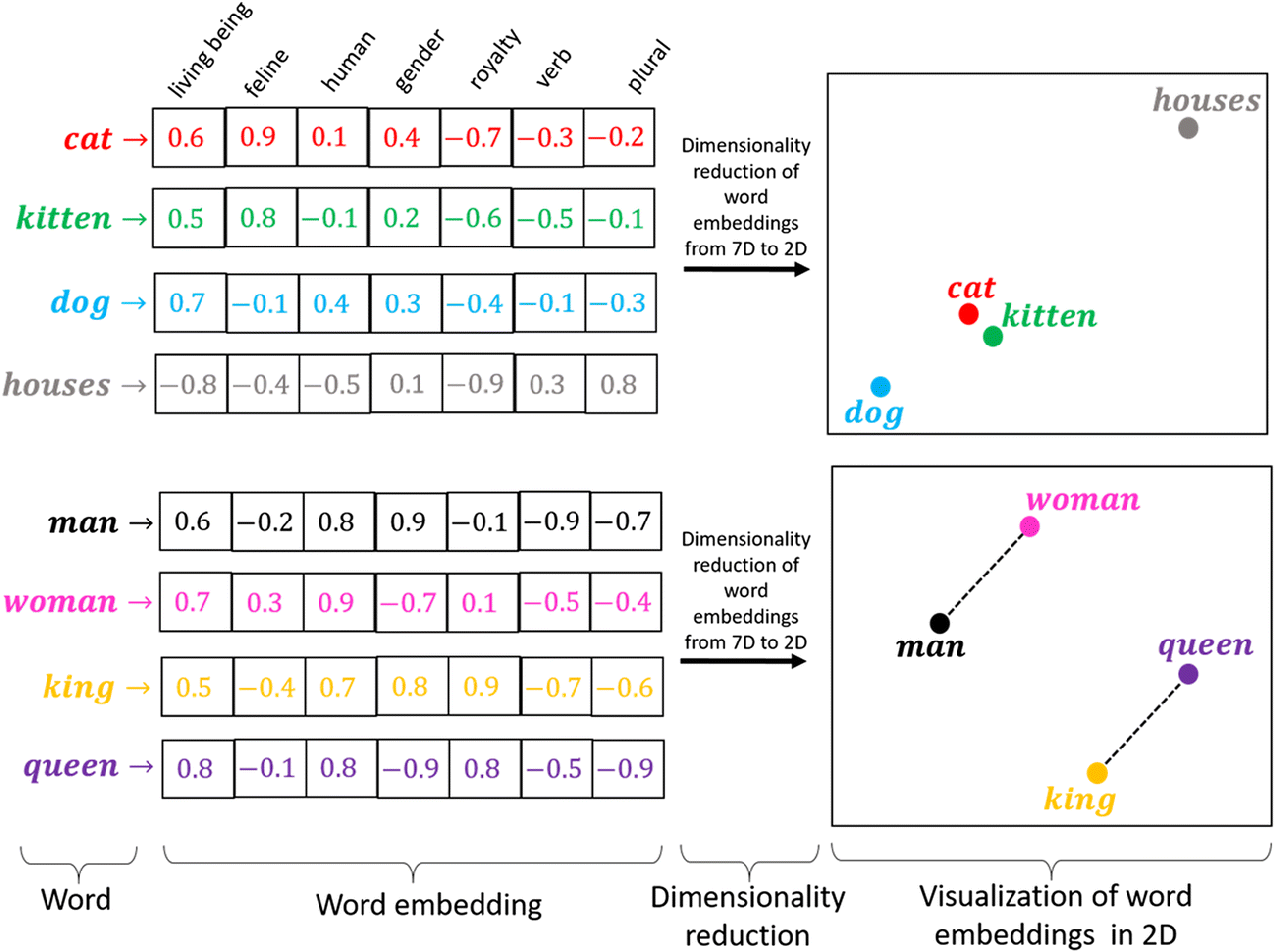
ChatGPT-User

`ChatGPT-User` is used by plugins in ChatGPT. This user-agent will only be used to take direct actions on behalf of ChatGPT users and is *not* used for crawling the web in any automatic fashion.

- User agent token: `ChatGPT-User`
- Full user-agent string: `Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko); compatible; ChatGPT-User/1.0; +https://openai.com/bot`

To allow plugins to access your site you can explicitly add the `ChatGPT-User` to your site's robots.txt:

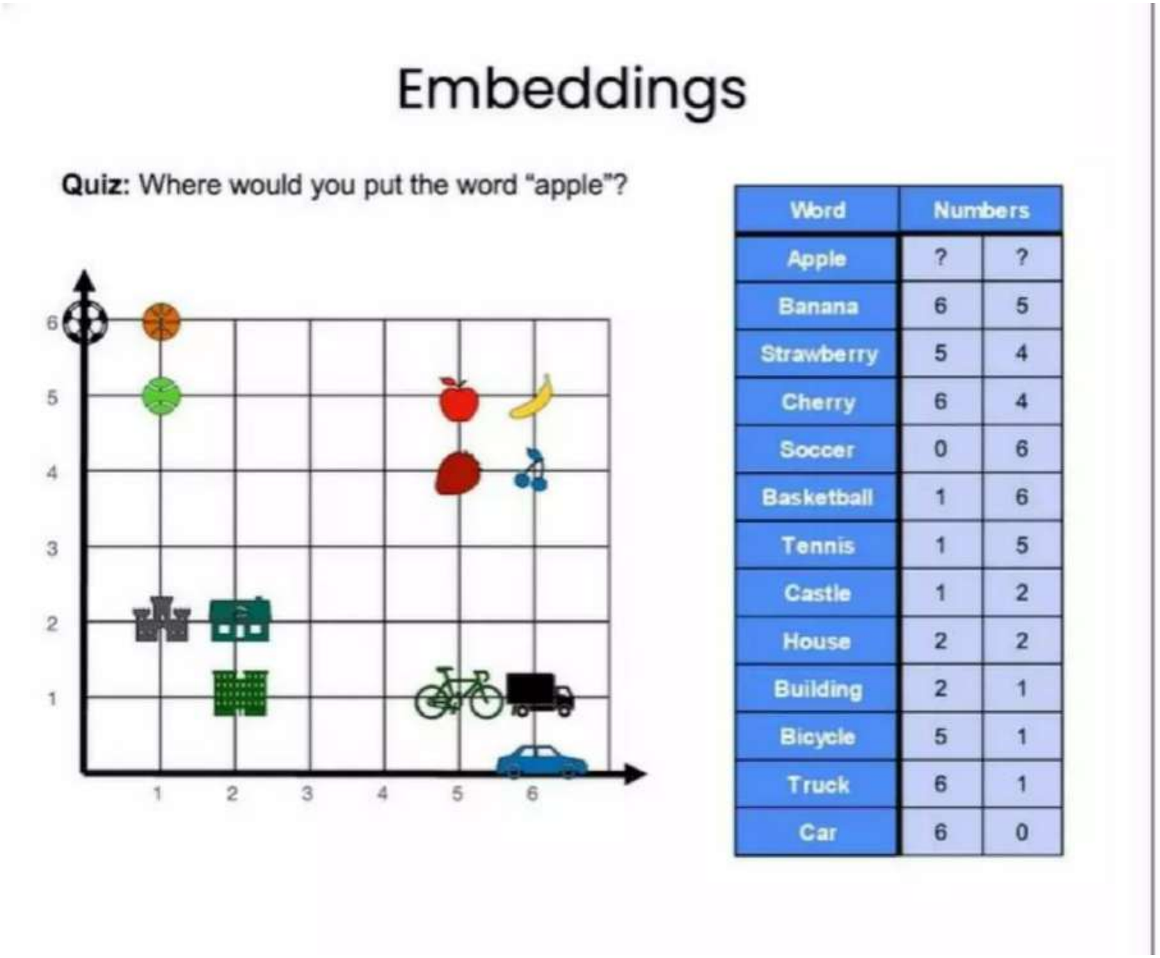
```
User-agent: ChatGPT-User
Disallow:
```



	living being	feline	human	gender	royalty	verb	plural
cat →	0.6	0.9	0.1	0.4	−0.7	−0.3	−0.2
kitten →	0.5	0.8	−0.1	0.2	−0.6	−0.5	−0.1
dog →	0.7	−0.1	0.4	0.3	−0.4	−0.1	−0.3
houses →	−0.8	−0.4	−0.5	0.1	−0.9	0.3	0.8

man →	0.6	−0.2	0.8	0.9	−0.1	−0.9	−0.7
woman →	0.7	0.3	0.9	−0.7	0.1	−0.5	−0.4
king →	0.5	−0.4	0.7	0.8	0.9	−0.7	−0.6
queen →	0.8	−0.1	0.8	−0.9	0.8	−0.5	−0.9

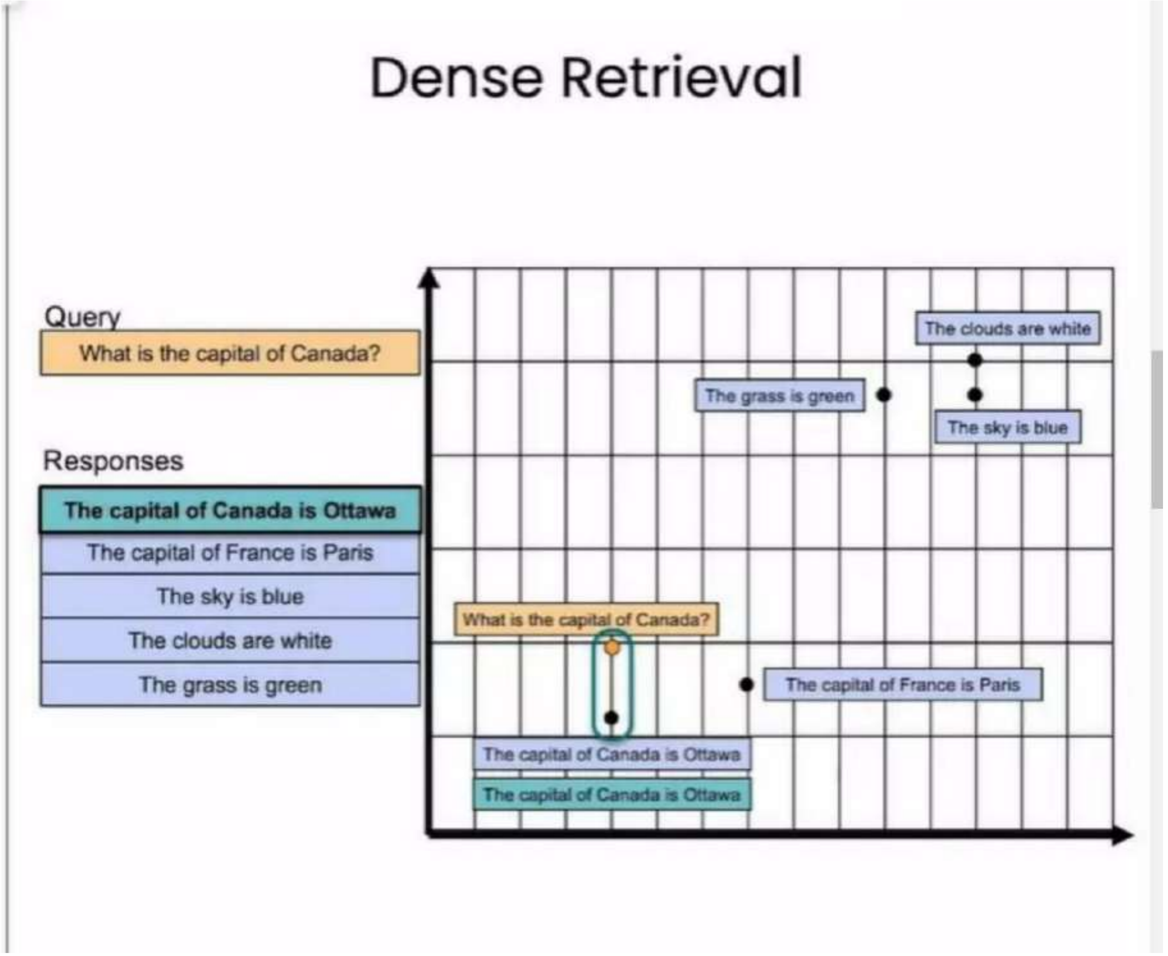
Dimensionality reduction of word embeddings from 7D to 2D
houses
cat
kitten
dog
Dimensionality reduction of word embeddings from 7D to 2D
woman
man
queen
king
Word
Word embedding
Dimensionality reduction
Visualization of word embeddings in 2D



Embeddings

Quiz: Where would you put the word “apple”?

Word	Numbers	
Apple	?	?
Banana	6	5
Strawberry	5	4
Cherry	6	4
Soccer	0	6
Basketball	1	6
Tennis	1	5
Castle	1	2
House	2	2
Building	2	1
Bicycle	5	1
Truck	6	1
Car	6	0



Dense Retrieval

Query

What is the capital of Canada?

Responses

The capital of Canada is Ottawa

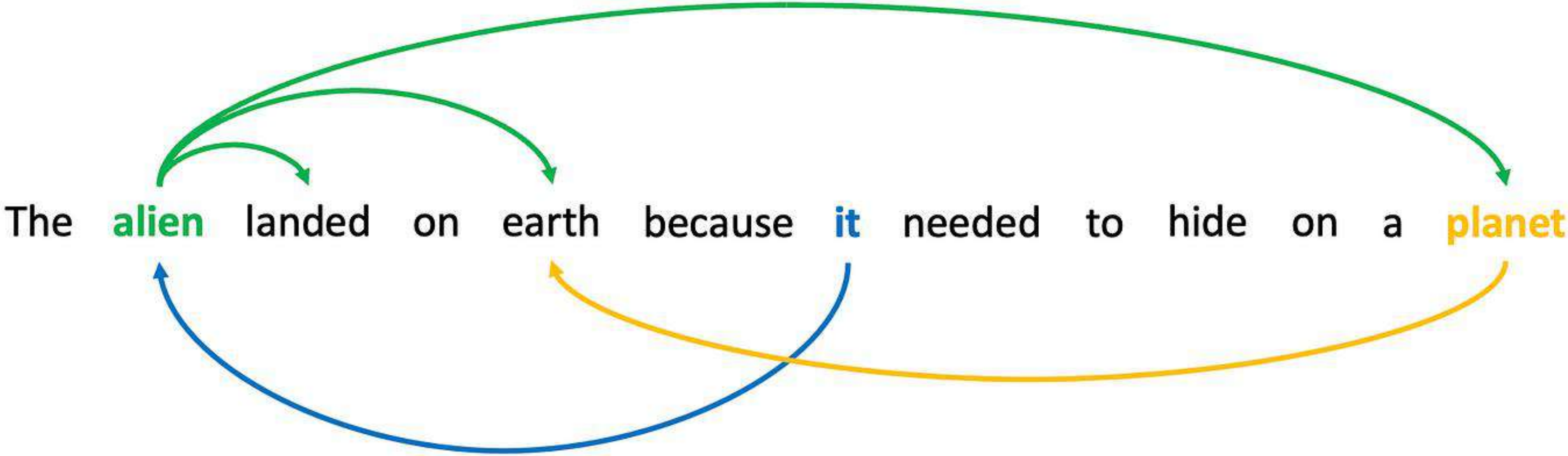
The capital of France is Paris

The sky is blue

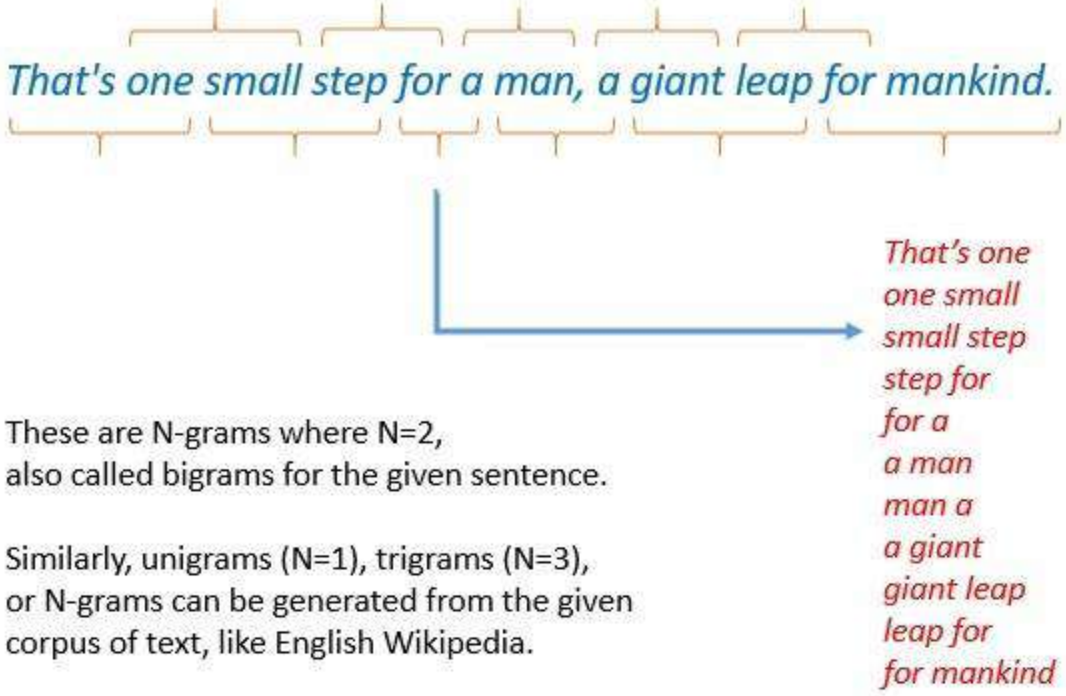
The clouds are white

The grass is green

The clouds are white
The grass is green
The sky is blue
What is the capital of Canada?
The capital of France is Paris
The capital of Canada is Ottawa
The capital of Canada is Ottawa



The **alien** landed on earth because **it** needed to hide on a **planet**



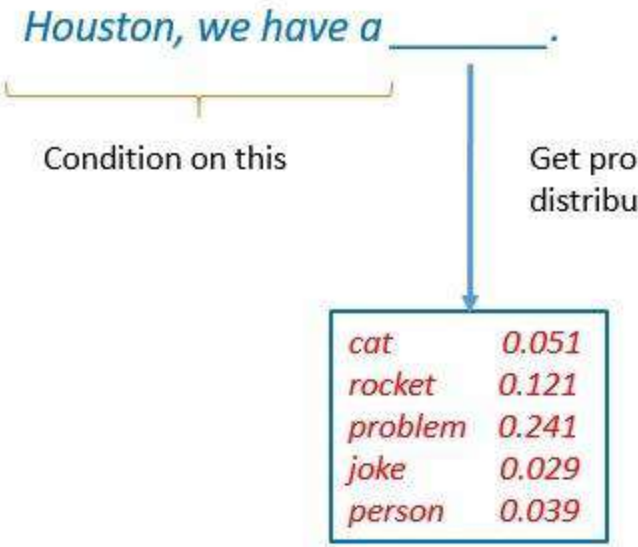
That's one small step for a man, a giant leap for mankind.

That's one
one small
small step
step for
for a
a man
man a
a giant
giant leap
leap for
for mankind

These are N-grams where N=2,
also called bigrams for the given sentence.

Similarly, unigrams (N=1), trigrams (N=3),
or N-grams can be generated from the given
corpus of text, like English Wikipedia.

Now, a language model based on statistics on frequency of n-grams can be used to predict the next word in a sequence:

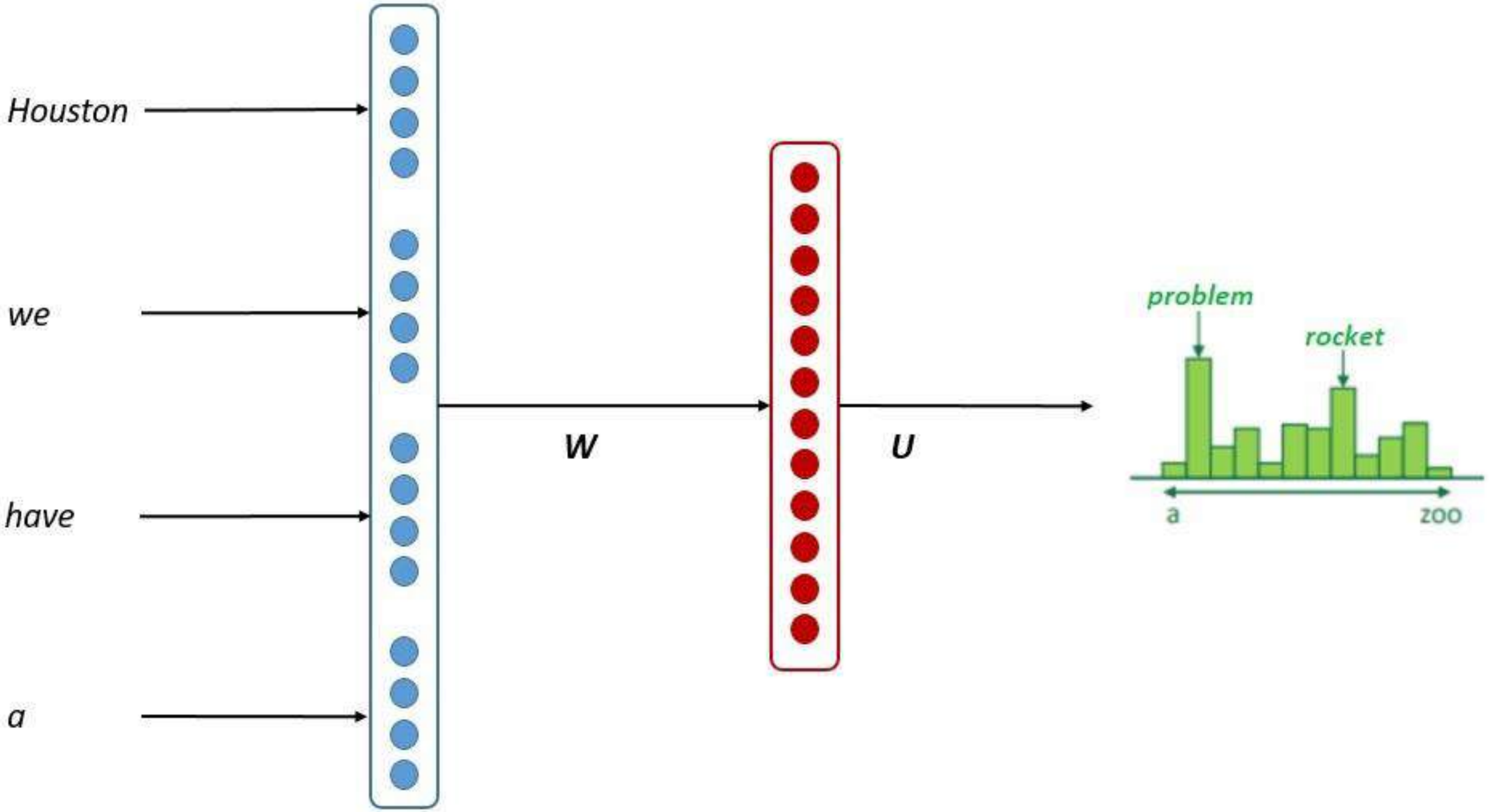


Houston, we have a ________.

Condition on this

Get probability distribution

cat	*0.051*
rocket	*0.121*
problem	*0.241*
joke	*0.029*
person	*0.039*



x^1, x^2, x^3, x^4

Input text sequence

$e = [e^1; e^2; e^3; e^4]$

Concatenated word embeddings

$h = f(We + b_1)$

Hidden layer of the neural network

$\hat{y} = softmax(Uh + b_2)$

Output probability distribution

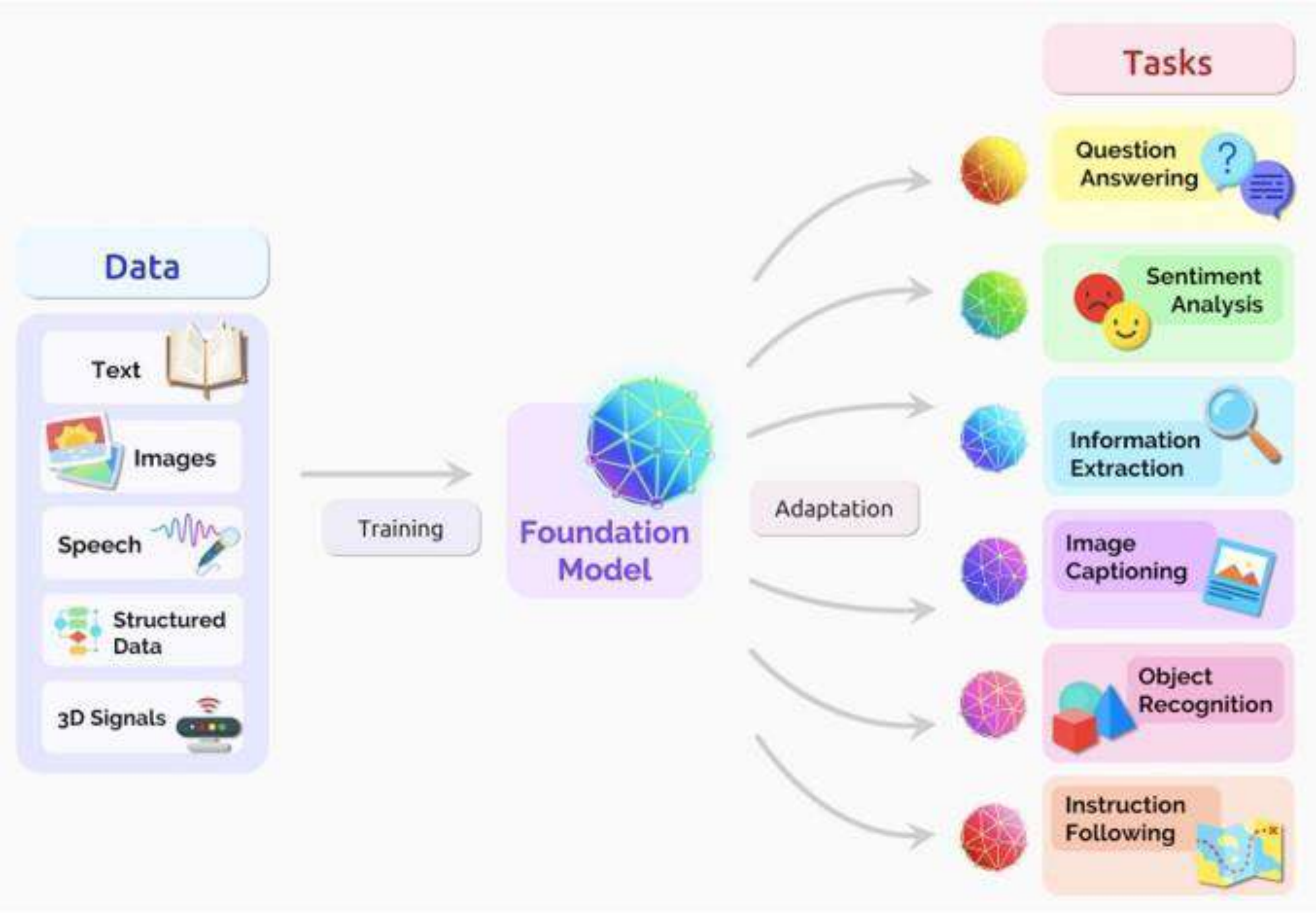
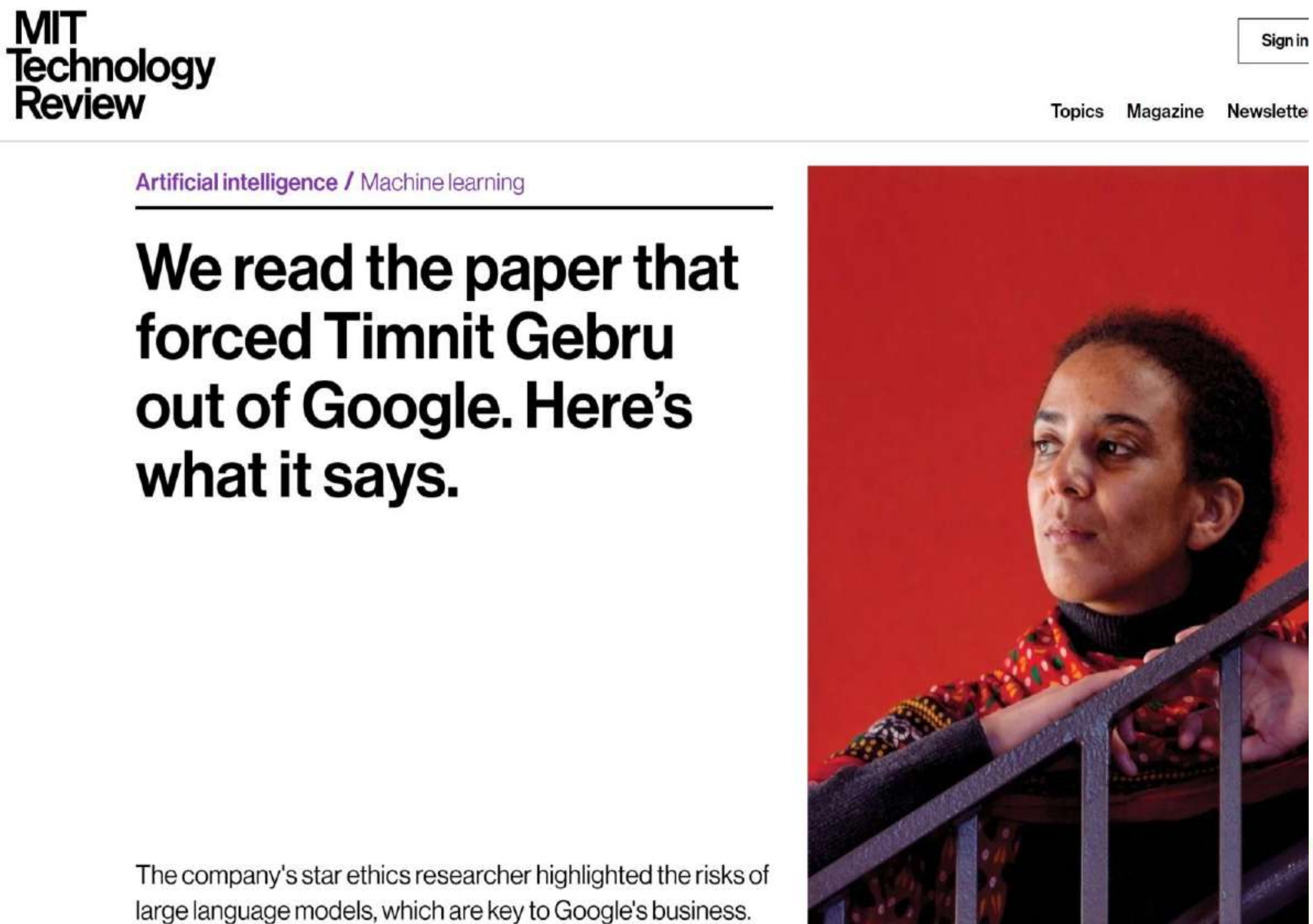


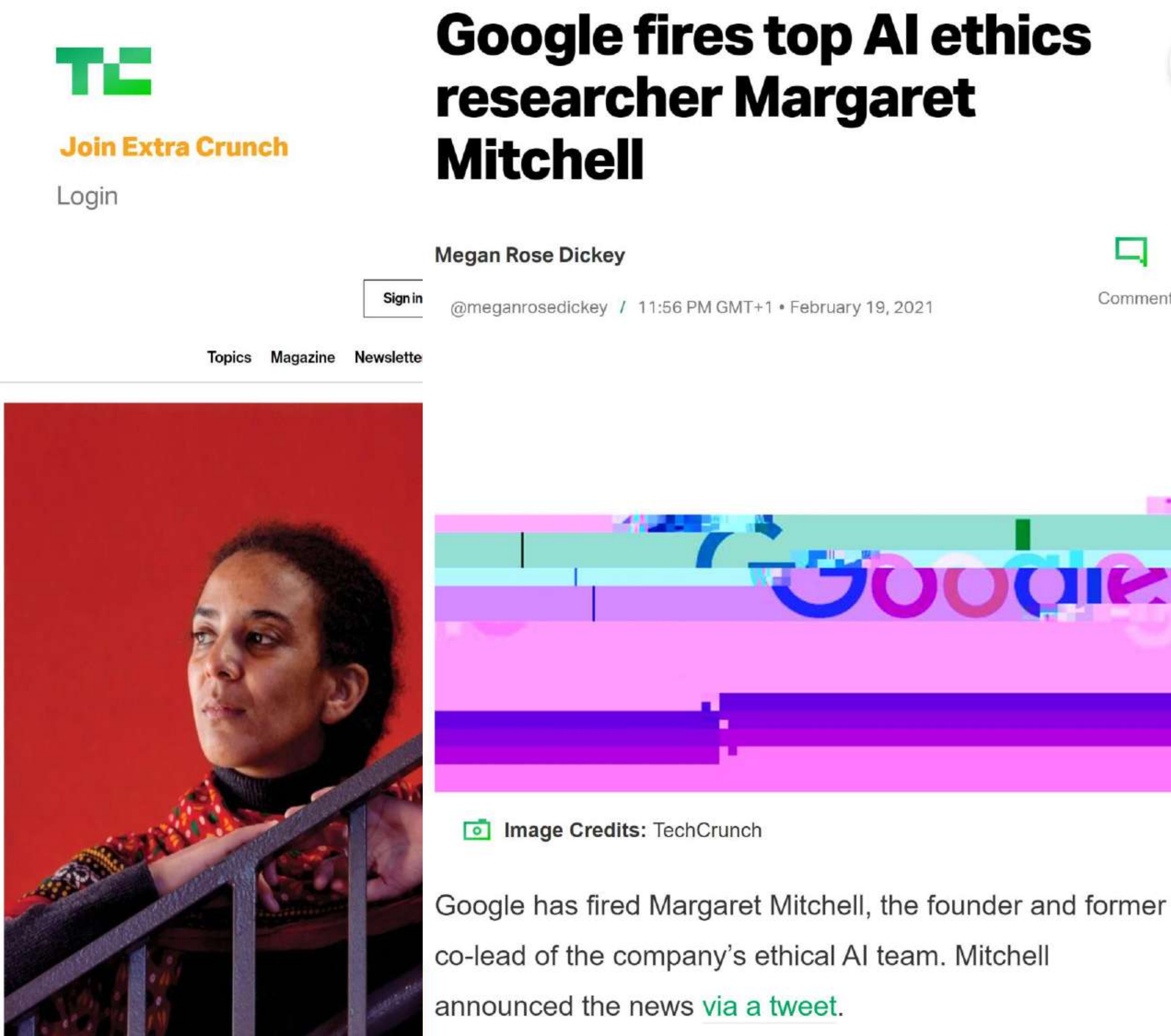
Image Credit: https://research.aimultiple.com/foundation-models/



Artificial intelligence / Machine learning

We read the paper that forced Timnit Gebru out of Google. Here's what it says.

The company's star ethics researcher highlighted the risks of large language models, which are key to Google's business.



Google fires top AI ethics researcher Margaret Mitchell

Megan Rose Dickey

@meganrosedickey / 11:56 PM GMT+1 • February 19, 2021

Image Credits: TechCrunch

Google has fired Margaret Mitchell, the founder and former co-lead of the company's ethical AI team. Mitchell announced the news via a tweet.

AIES READING GROUP

On the Dangers of Stochastic Parrots:

Can Language Models Be Too Big?

24 Feb 2021
3-4.30pm GMT

UNFATHOMABLE TRAINING DATA

(unfathomable : incapable of being fully explored or understood)

- Size Doesn’t Guarantee Diversity
- Static Data/Changing Social Views
- Encoding Bias
- Curation, Documentation & Accountability

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 Emily M Bender, **Timnit Gebru**, Angelina McMillan-Major, **Shmargaret Shmitchell**

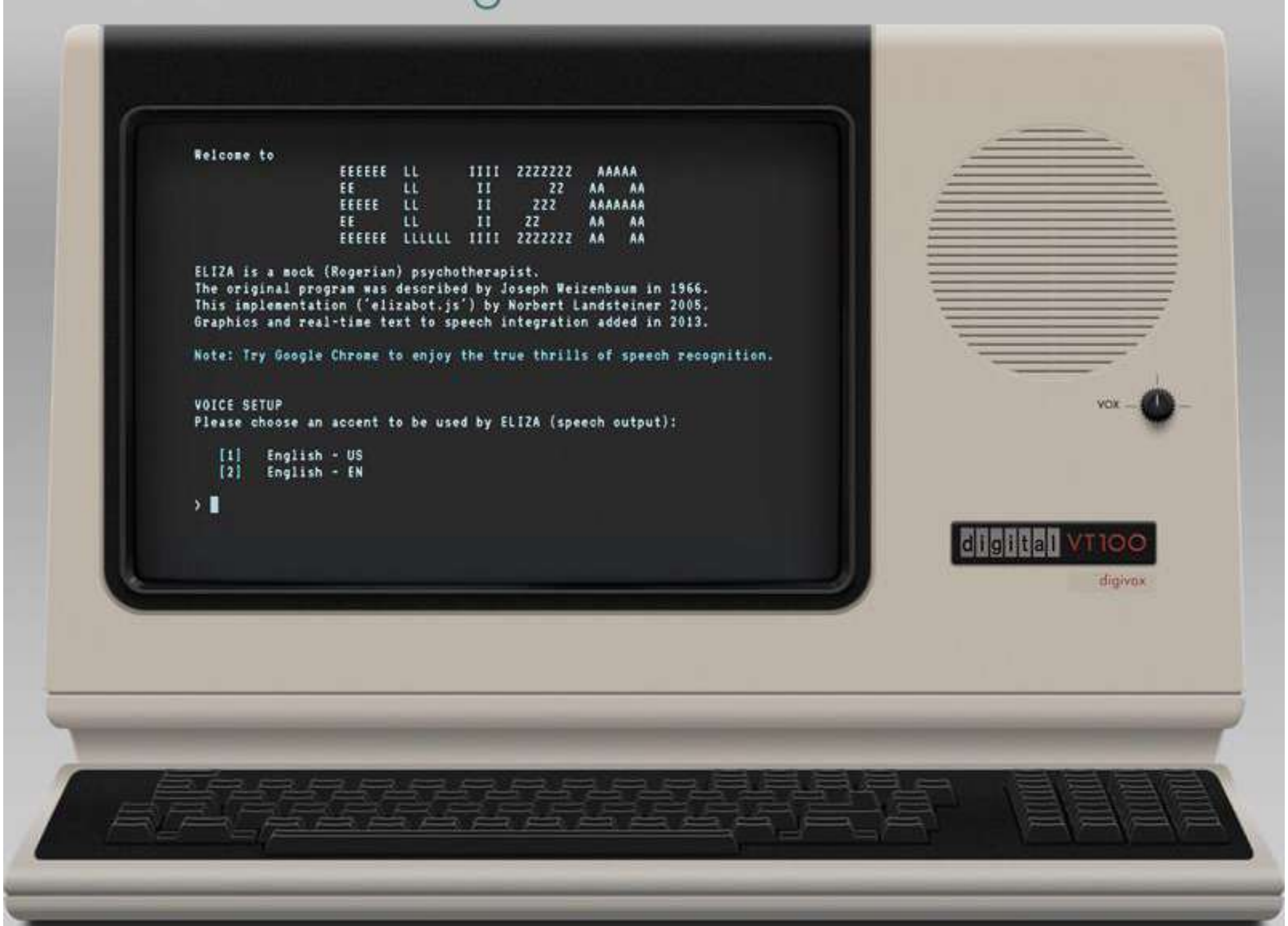
“Any product development that involves **operationalizing definitions** around such **shifting topics into algorithms** is necessarily **political** (whether or not developers choose the path of **maintaining the *status quo ante***)”

Manipulation of users

“If a large LM, endowed with hundreds of billions of parameters and trained on a very large dataset, can **manipulate linguistic form** well enough **to cheat its way through tests** meant to require **language understanding,** have we learned anything of value about how to build machine language understanding or have we been led down the garden path?” We say seemingly coherent because **coherence is in fact in the eye of the beholder**. Our human understanding of coherence derives from our ability **to recognize interlocutors’ beliefs** [30, 31] and **intentions** [23, 33] within **context** [32]. As such, human communication relies on the **interpretation of implicit meaning conveyed between individuals**. The fact that human-human communication is a **jointly constructed activity** [29, 128] is most clearly true in co-situated spoken or signed communication

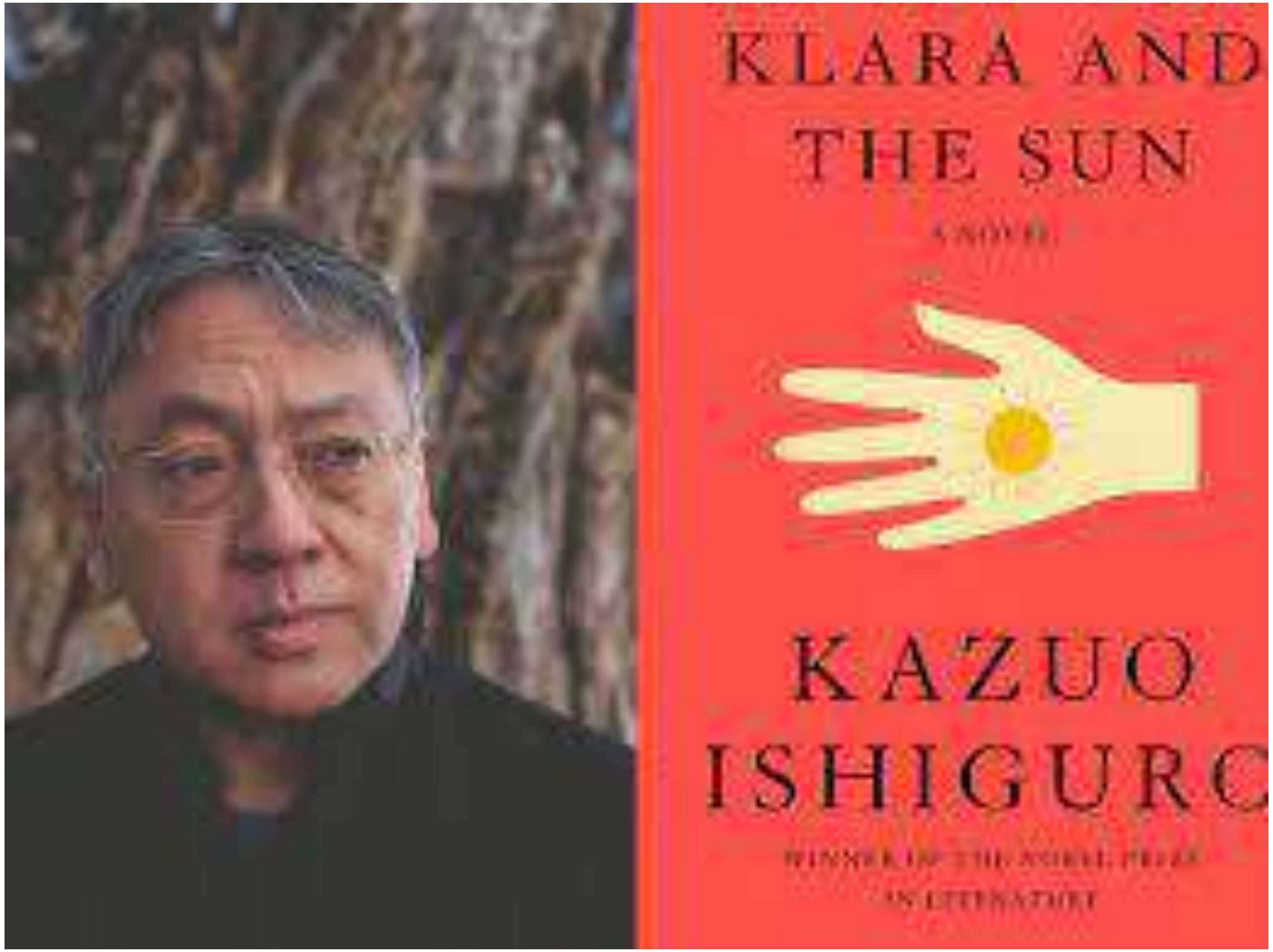
Text generated by an LM is **not grounded in communicative intent**, any model of the world, or any **model** of the **reader’s state of mind**. It can’t have been, because the **training data never included sharing thoughts with a listener,** nor does the machine have the ability to do that-

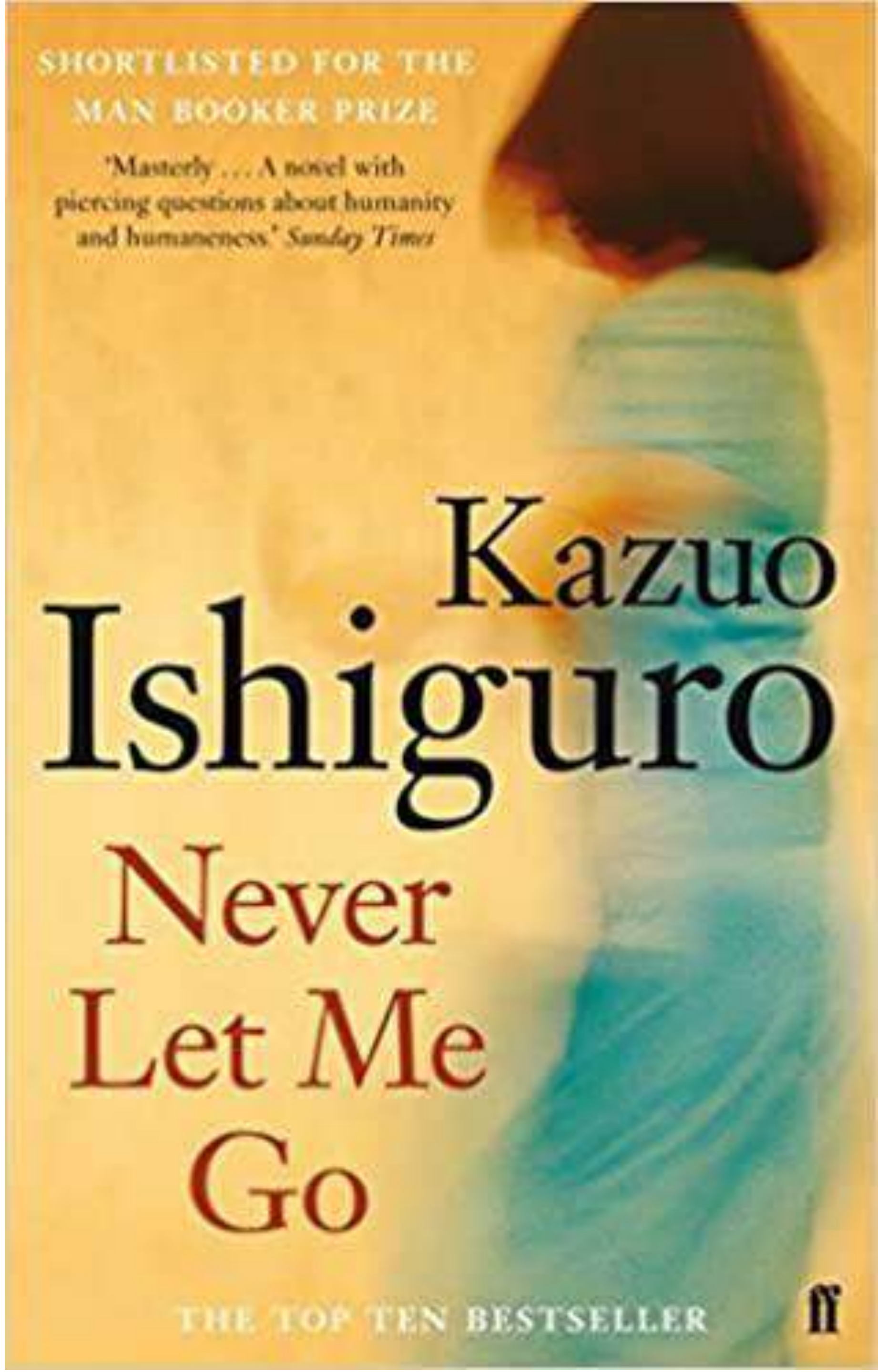
The problem is, **if one side of the communication does not have meaning**, then the **comprehension of the implicit meaning is an illusion arising from our singular human understanding of languag**e (independent of the model). Contrary to how it may seem when we observe its output, **an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data**, according to probabilistic information about how they combine, but without any reference to meaning: a <u>**stochastic parrot**</u>.



```
Welcome to
                  EEEEEE  LL      IIII  2222222   AAAAA
                  EE      LL       II       22   AA   AA
                  EEEEE   LL       II     222    AAAAAAA
                  EE      LL       II    22      AA   AA
                  EEEEEE  LLLLLL  IIII  2222222  AA   AA

ELIZA is a mock (Rogerian) psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation ('elizabot.js') by Norbert Landsteiner 2005.
Graphics and real-time text to speech integration added in 2013.

Note: Try Google Chrome to enjoy the true thrills of speech recognition.


VOICE SETUP
Please choose an accent to be used by ELIZA (speech output):

   [1]   English - US
   [2]   English - EN

> █
```

od
mit


KLARA AND
THE SUN
A NOVEL
KAZUO
ISHIGURO
WINNER OF THE NOBEL PRIZE
IN LITERATURE



SHORTLISTED FOR THE MAN BOOKER PRIZE

'Masterly . . . A novel with piercing questions about humanity and humaneness.' *Sunday Times*

Kazuo Ishiguro

Never Let Me Go

THE TOP TEN BESTSELLER

ff

BENEFITS OF CHAT GPT

CONTENT GENERATION

ChatGPT can be used to generate content such as summaries, reviews, and essays. It is also useful as a virtual assistant for students, helping them turn out content for assignments.

VIRTUAL TUTORING

ChatGPT can act as a virtual tutor, providing one-on-one instruction answering questions in real time. The chatbot is responsive with no waiting times and a quick response with precise feedback.

LANGUAGE LEARNING

ChatGPT can also be used to help students learn a new language. . It can also check student essays for grammar, vocabulary, and coherence.

WIDE RANGE OF RESOURCES

ChatGPT can provide you with access to a wide range of resources, including study materials, practice exams, and educational videos. This can help you to learn more effectively and efficiently.

POSITIVE , NEGATIVE EFFECTS

ONE OF THE KEY BENEFITS OF CHATGPT IS THAT IT IS AVAILABLE 24/7 TO PROVIDE SUPPORT AND GUIDANCE TO STUDENTS.

24/7 ACCESS

CHATBOTS ARE BEING USED TO PROVIDE ADVICE TO STUDENTS ON ACADEMIC ISSUES TO MAKE SOME VITAL DECISIONS ACA-DEMIC ACTIVITIES.

ADVISORY

THE ABILITY OF TEACHER-S TO UPLOAD INFORMATION ABOUT A SPECIFIC SUBJECT TO AN ONLINE PLATFORM FOR EASY ACCESS BY AUTHORIZED STUDENTS.

INTEGRATION OF CONTENTS

CHATBOTS THAT CAN BE USED TO DELIVER ADMINISTRATIVE TASKS IN EDUCATIONAL INSTITUTIONS.

ADMINISTRATION

BUILDING A CHATBOT SYSTEM IS A CONTINUOUS PROCESS THAT NEED CONSISTENT SUPERVISION AND MAINTENANCE, WHICH CAN BE DIFFICULT.

SUPERVISION AND MAINTENANCE

PLURALITY OF APPROACHES, TRUST AND TRANSPARENCY, PRIVACY, AND AGENT PERSONA.

ETHICAL ISSUE

IF STUDENTS HAVE NEGATIVE PERCEPTIONS OF CHATBOT TECHNOLOGY, THEY WILL BE HESITANT TO ADOPT AND USE THE TECHNOLOGY.

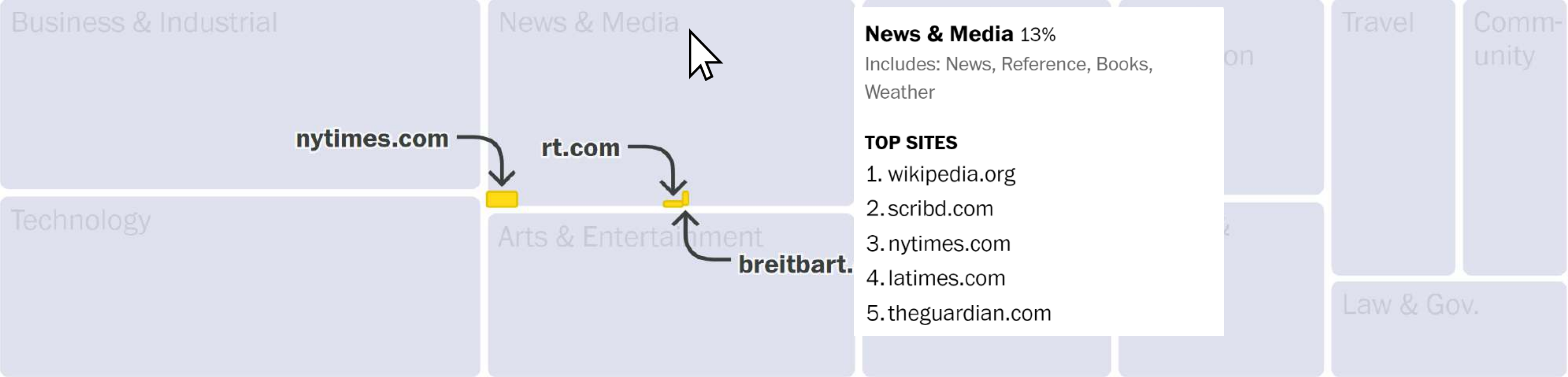
USER ATTITUDE ISSUE

WHILE CHATGPT TEXT IS CONSIDERED A DERIVATIVE WORK OF TRAINED DATA, THE COPYRIGHT STILL BELONGS TO THE ORIGINAL AUTHOR.

EVALUATION AND PLAGERISM

Problemi legati a ChatGPT e LLM

- Allucinazioni,
- Copyright,
- Speech acts,
- …



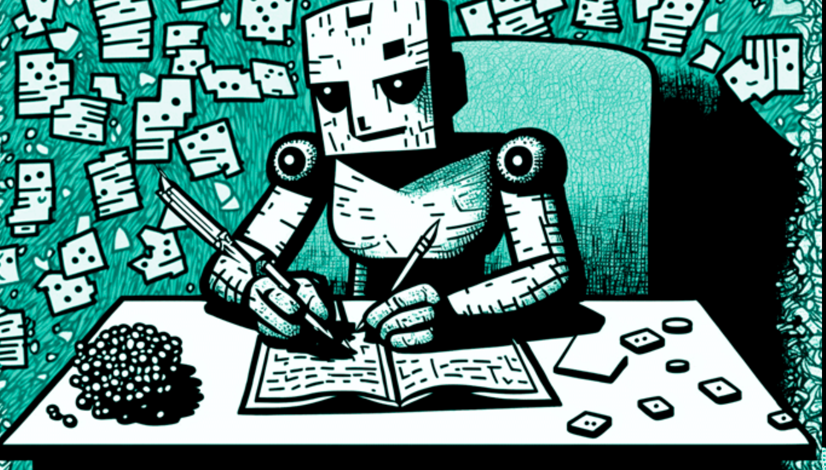
News & Media 13%

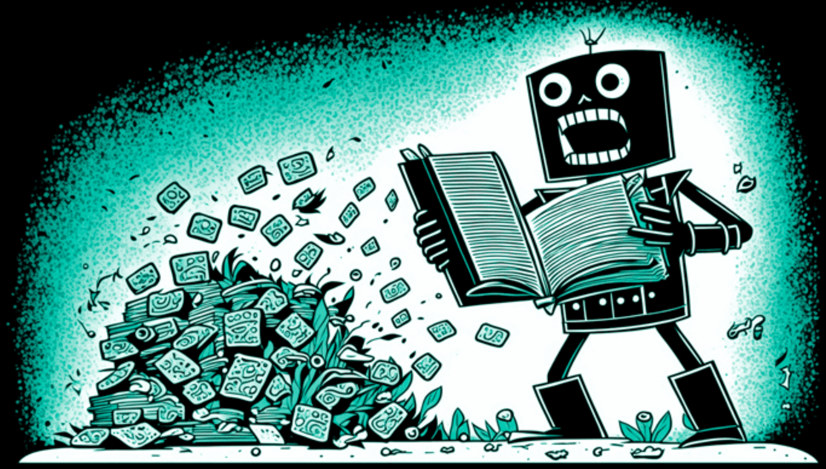
Includes: News, Reference, Books, Weather

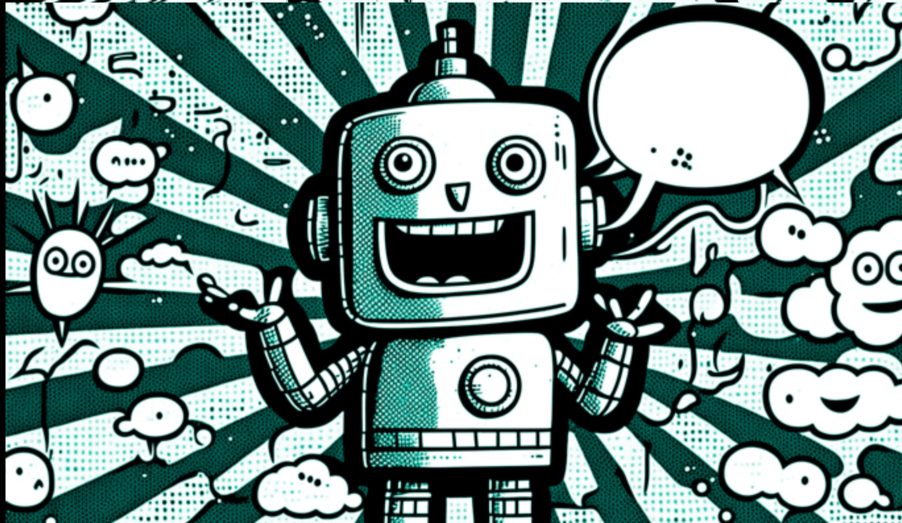
TOP SITES

1. wikipedia.org
2. scribd.com
3. nytimes.com
4. latimes.com
5. theguardian.com

Meanwhile, we found several media outlets that rank low on NewsGuard’s independent scale for trustworthiness: RT.com No. 65, the Russian state-backed propaganda site; breitbart.com No. 159, a well-known source for far-right news and opinion; and vdare.com No. 993, an anti-immigration site that has been associated with white supremacy.

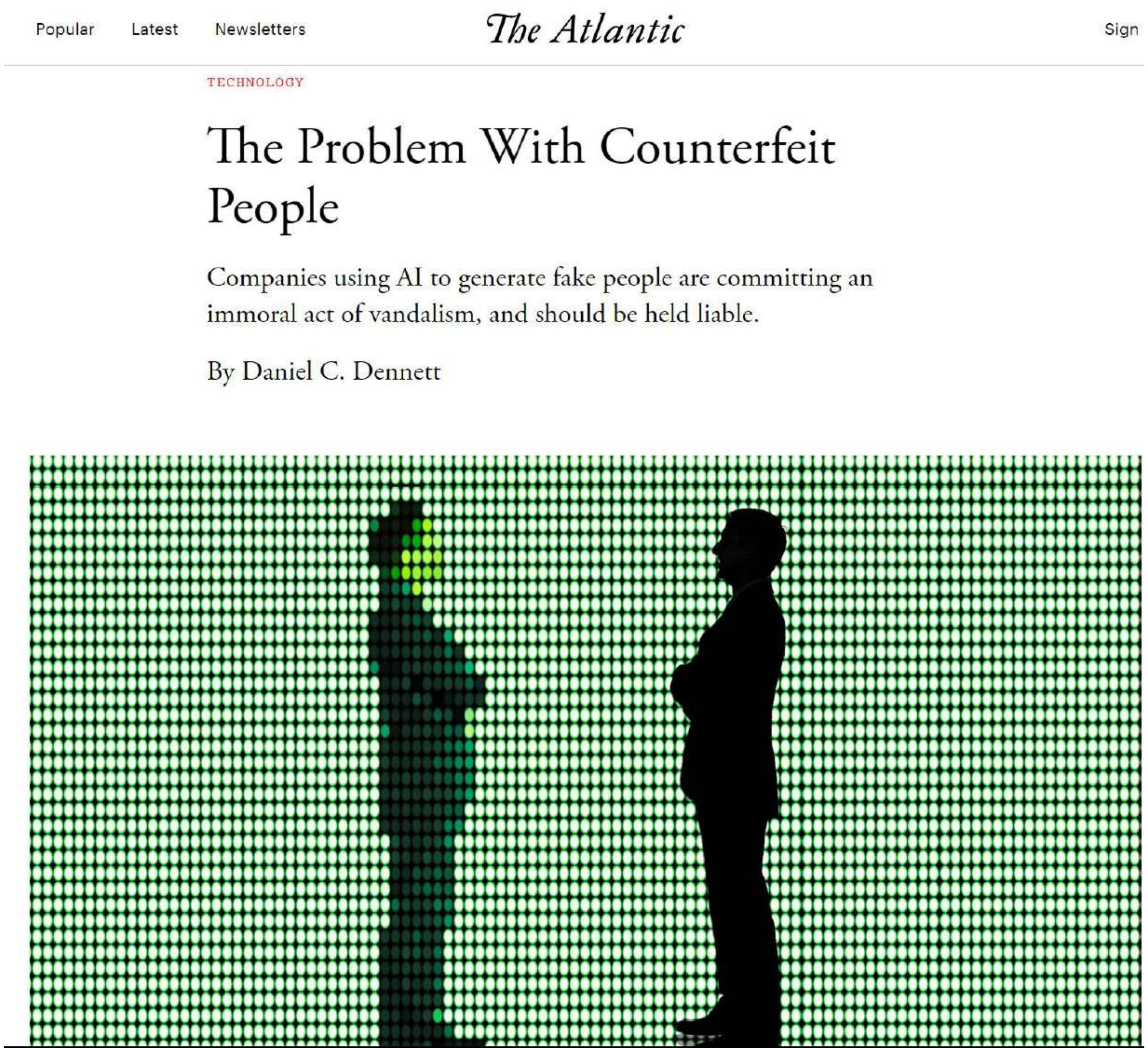






Artificial General Intelligence(AGI)

>> The Future of Intelligent Machines



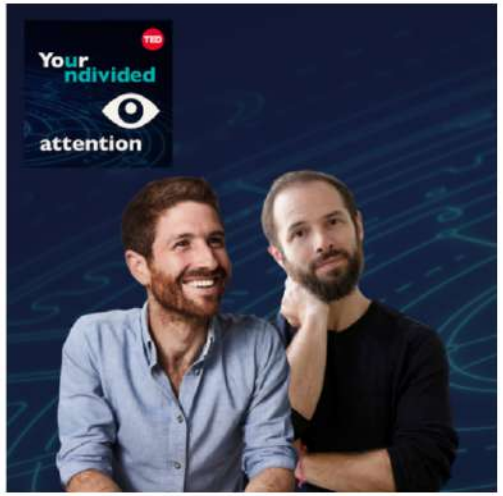
TECHNOLOGY

The Problem With Counterfeit People

Companies using AI to generate fake people are committing an immoral act of vandalism, and should be held liable.

By Daniel C. Dennett

See All Episodes



Synthetic Humanity: AI & What's At Stake

February 16, 2023

It may seem like the rise of artificial intelligence, and increasingly powerful large language models you may have heard of, is moving really fast… and it IS.

But what's coming next is when we enter synthetic relationships with AI that could come to feel just as real and important as our human relationships… And perhaps even more so.

SHARE EPISODE:

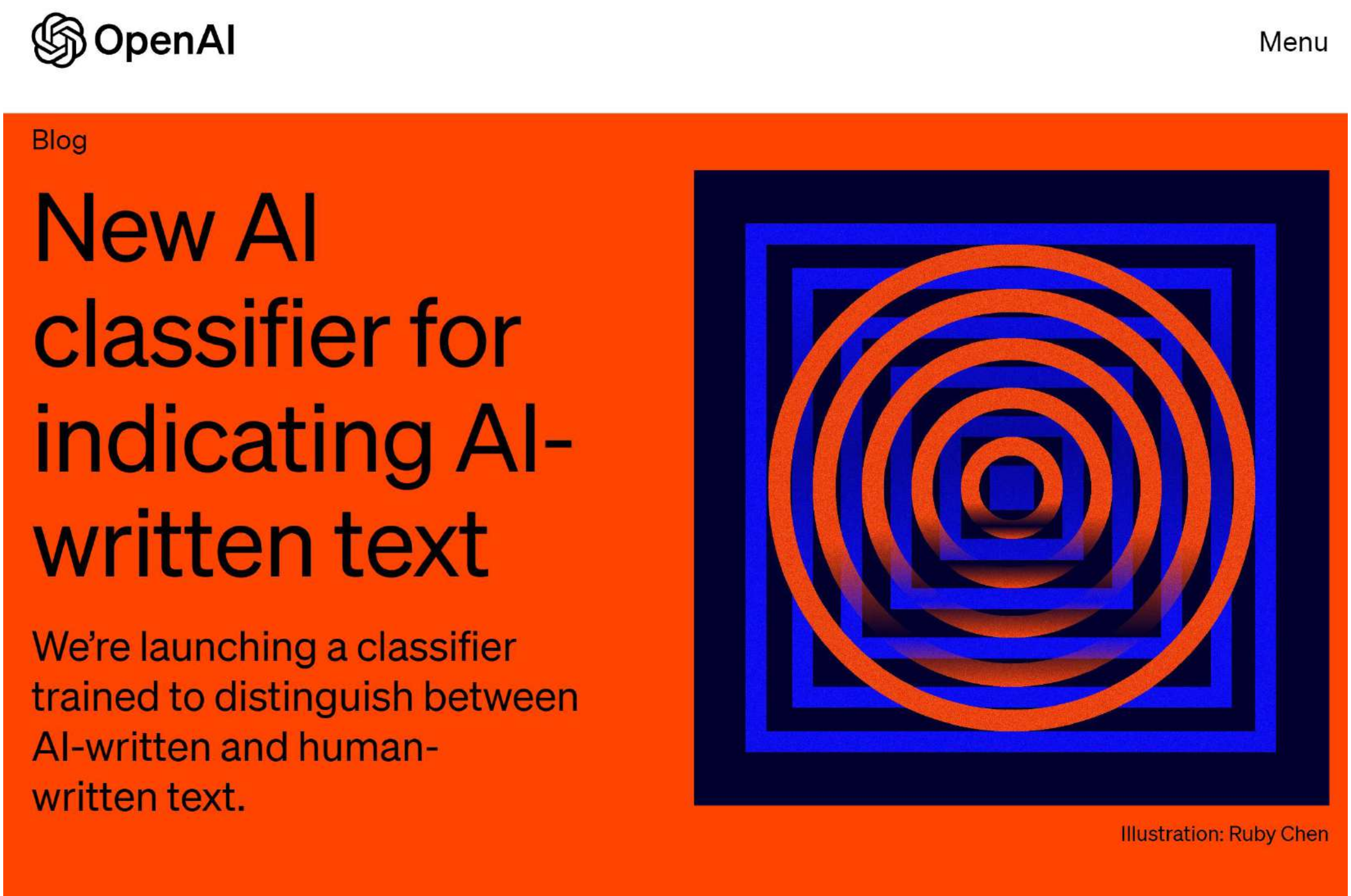
MY BLONDE GF

The experience of being deepfaked for pornography.

Watch now ➔

The Guardian
DOCUMENTARIES

BFI NETWORK doc society OKRE TALENTS TYKE FILMS



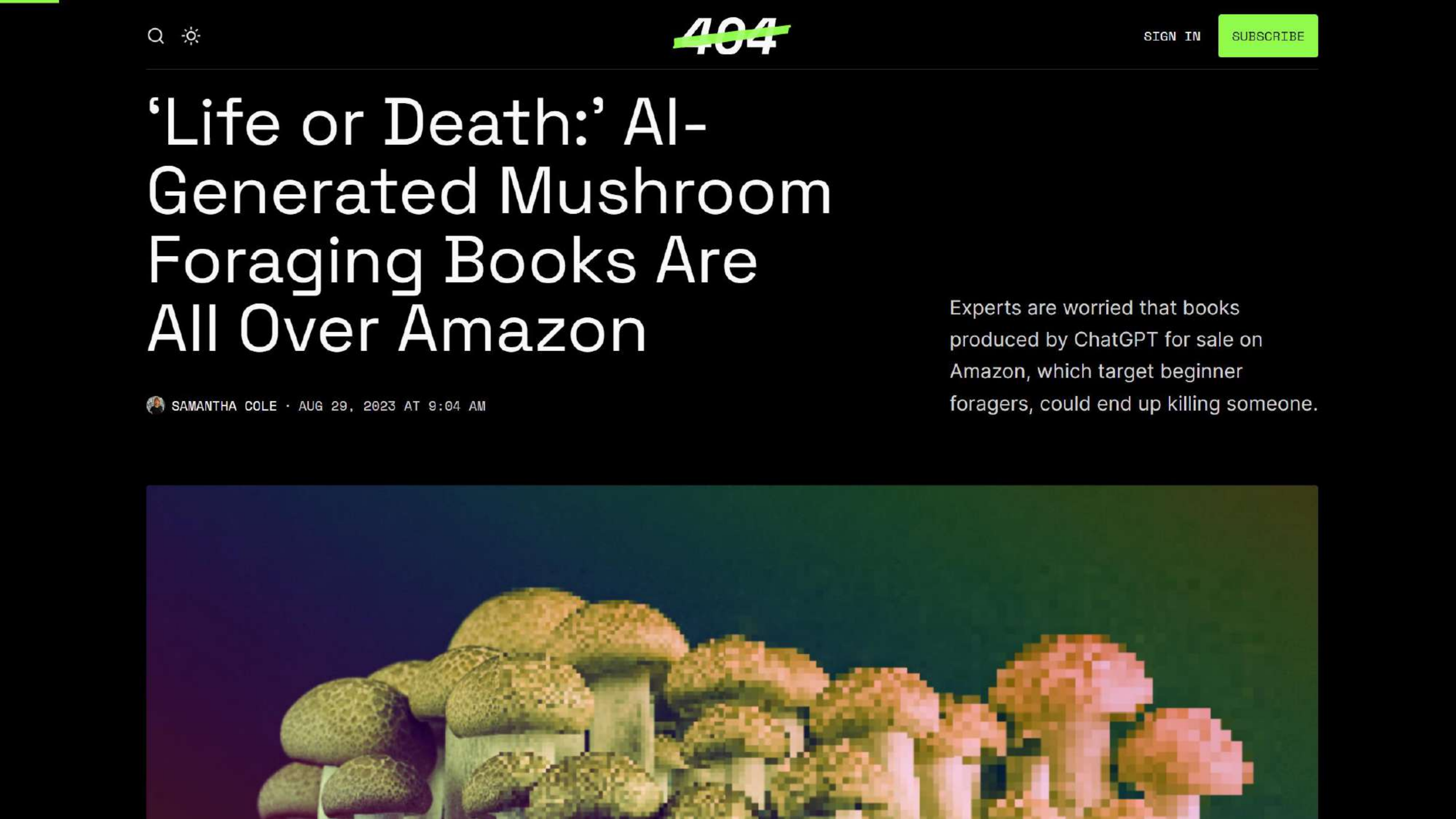
New AI classifier for indicating AI-written text

We're launching a classifier trained to distinguish between AI-written and human-written text.

Illustration: Ruby Chen

The New York Times
EVERYTHING YOU NEED TO KNOW FOR A MEMORABLE TRIP
NEW YORK CITY
TRAVEL GUIDE 2023
ADAM NEAL
ANTHONY TURNER
FAROE ISLANDS
TRAVEL GUIDE 2023
EVERYTHING YOU MUST KNOW BEFORE PLANNING A TRIP TO FAROE ISLANDS
MALTA
TRAVEL GUIDE 2023
EVERYTHING YOU NEED TO KNOW BEFORE PLANNING A TRIP TO MALTA
MADEIRA
TRAVEL GUIDE 2023
EVERYTHING YOU MUST KNOW BEFORE PLANNING A TRIP TO MADEIRA
DOLOMITES
TRAVEL GUIDE 2023
EVERYTHING YOU MUST KNOW BEFORE PLANNING A TRIP TO THE DOLOMITES
BORA BORA
TRAVEL GUIDE 2023
EVERYTHING YOU MUST KNOW BEFORE PLANNING A TRIP TO BORA BORA
AMALFI COAST
Travel Guide 2023-2024
A GUIDE FOR FIRST-TIME TRAVELERS TO THE MOST BEAUTIFUL COAST IN ITALY
Cathryn Swartz
PHILIPPINES TRAVEL GUIDE
June, 2023
Exploring the Archipelago A Comprehensive Journey Through the Riches of the Philippines
UPDATED
NORWAY
TRAVEL GUIDE 2023
CHRIS GARRAWAY
KOS
Travel Guide 2023
NATE GRANO
MALLORCA
Travel Guide 2023
NATE GRANO
COSTA RICA
Travel Guide 2023
by
Nate Grano
GREAT BRITAIN
TRAVEL GUIDE
YELLOWSTONE
NATIONAL PARK
TRAVEL GUIDE 2023-2024

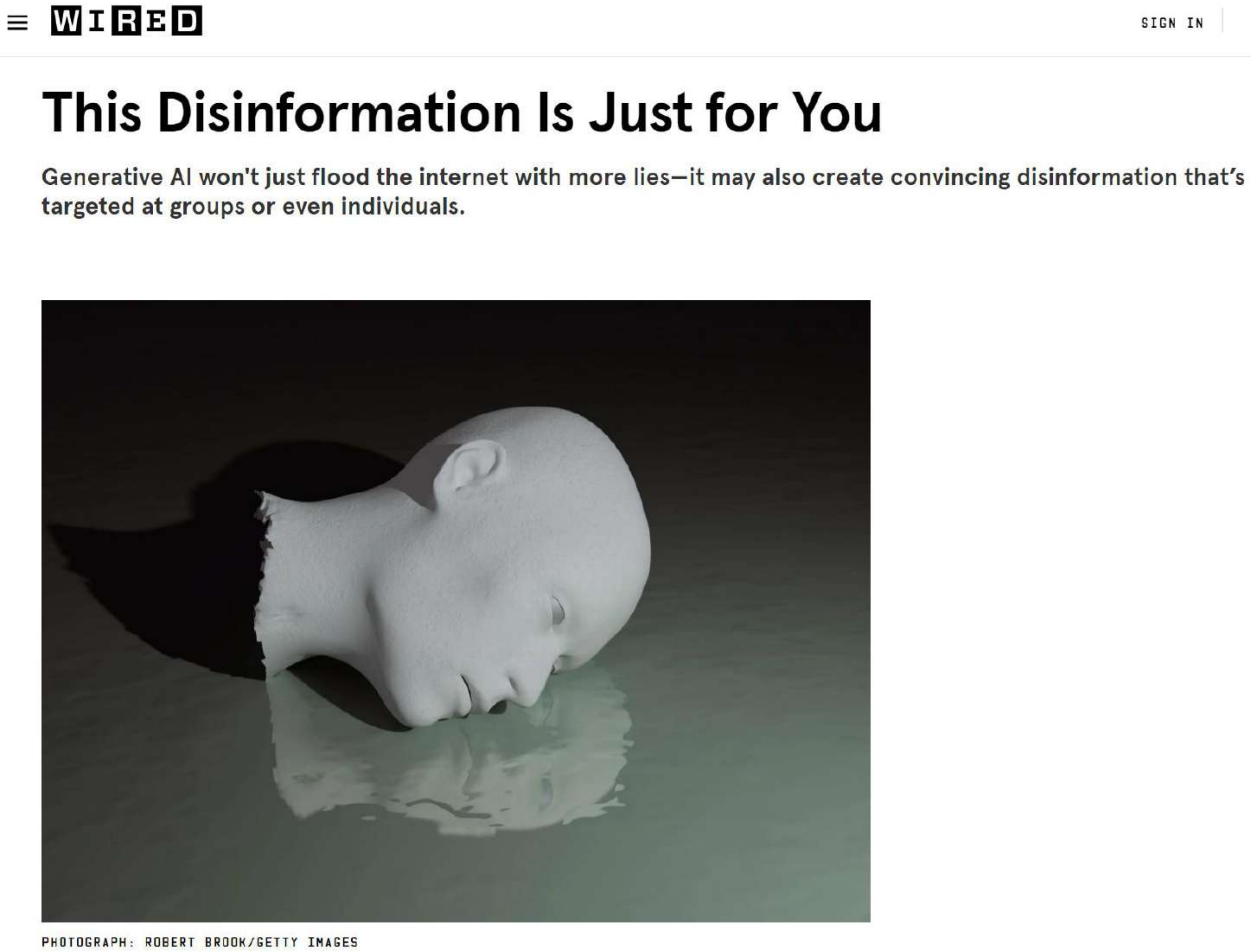
Shoddy guidebooks, promoted with deceptive reviews, have flooded Amazon in recent months.



‘Life or Death:’ AI-Generated Mushroom Foraging Books Are All Over Amazon

SAMANTHA COLE · AUG 29, 2023 AT 9:04 AM

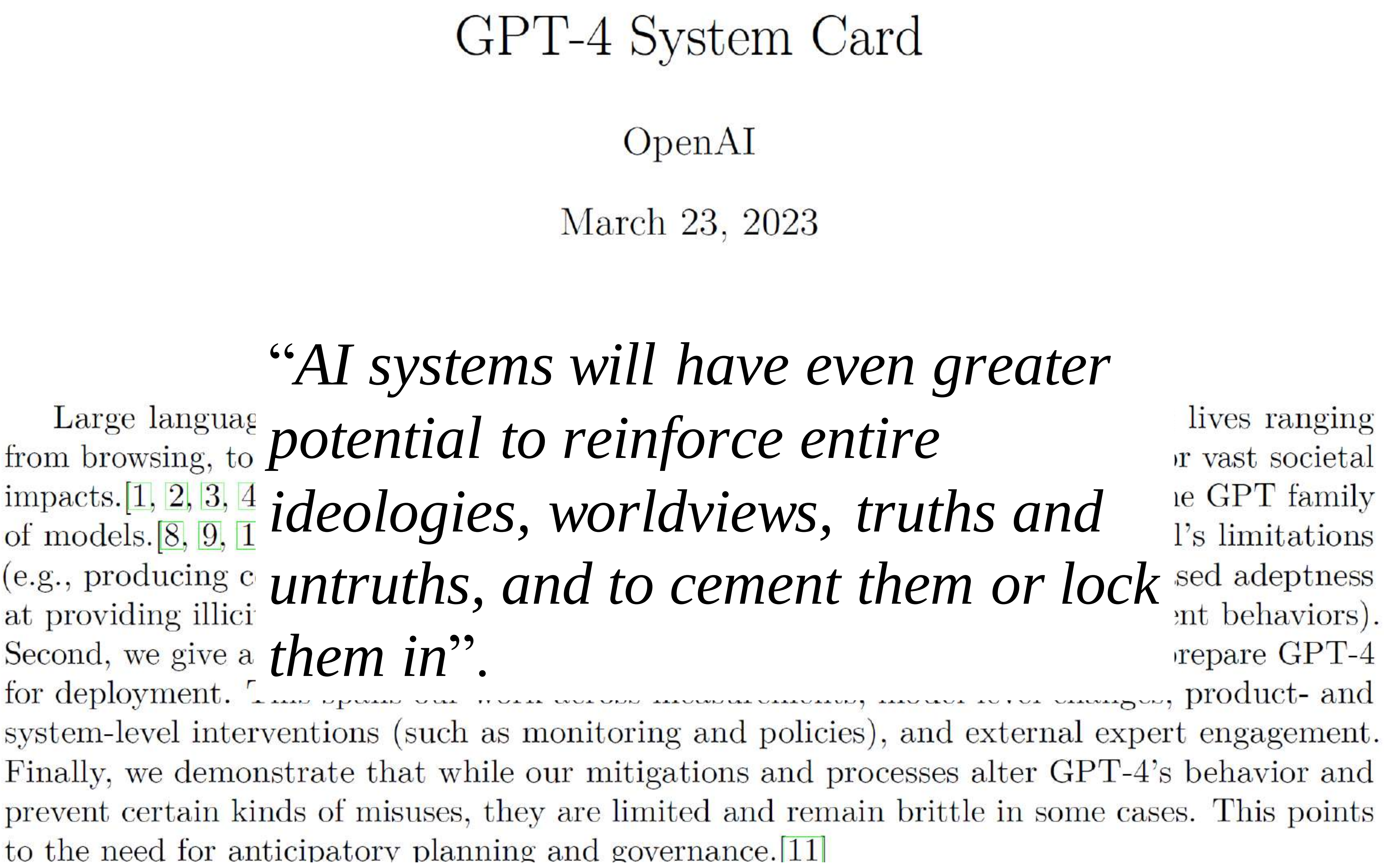
Experts are worried that books produced by ChatGPT for sale on Amazon, which target beginner foragers, could end up killing someone.



This Disinformation Is Just for You

Generative AI won't just flood the internet with more lies—it may also create convincing disinformation that’s targeted at groups or even individuals.

PHOTOGRAPH: ROBERT BROOK/GETTY IMAGES



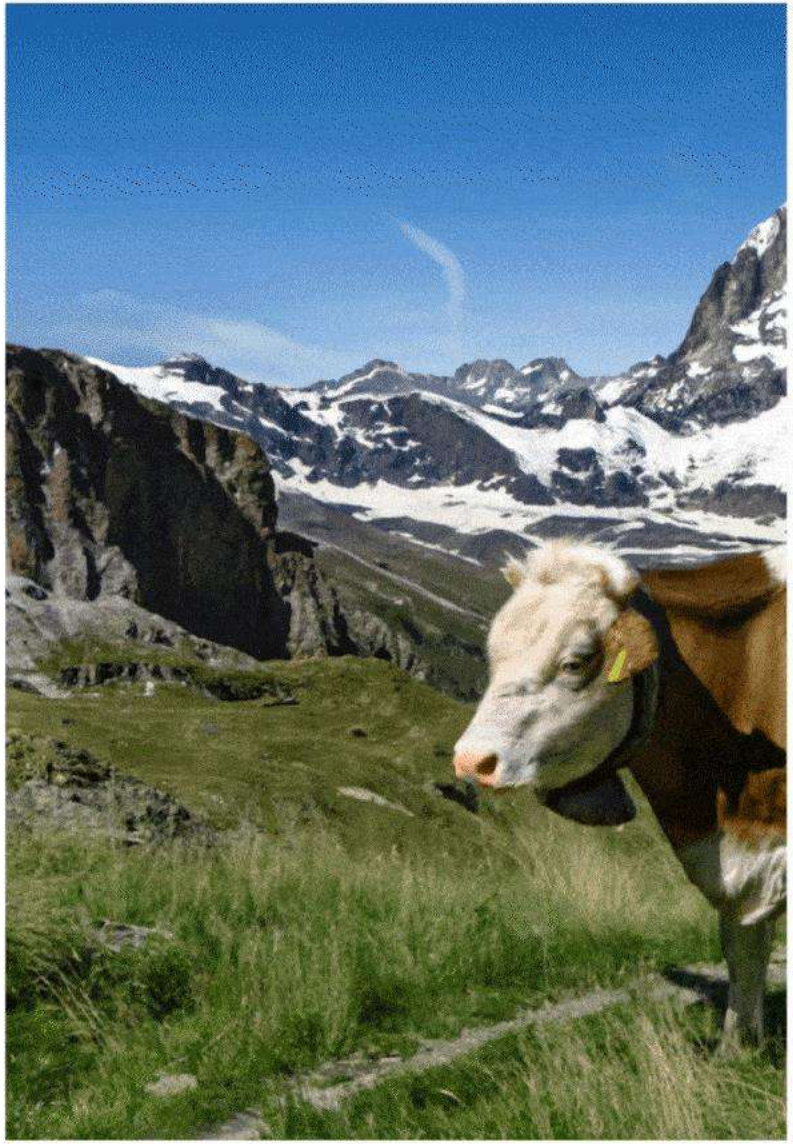
GPT-4 System Card

OpenAI

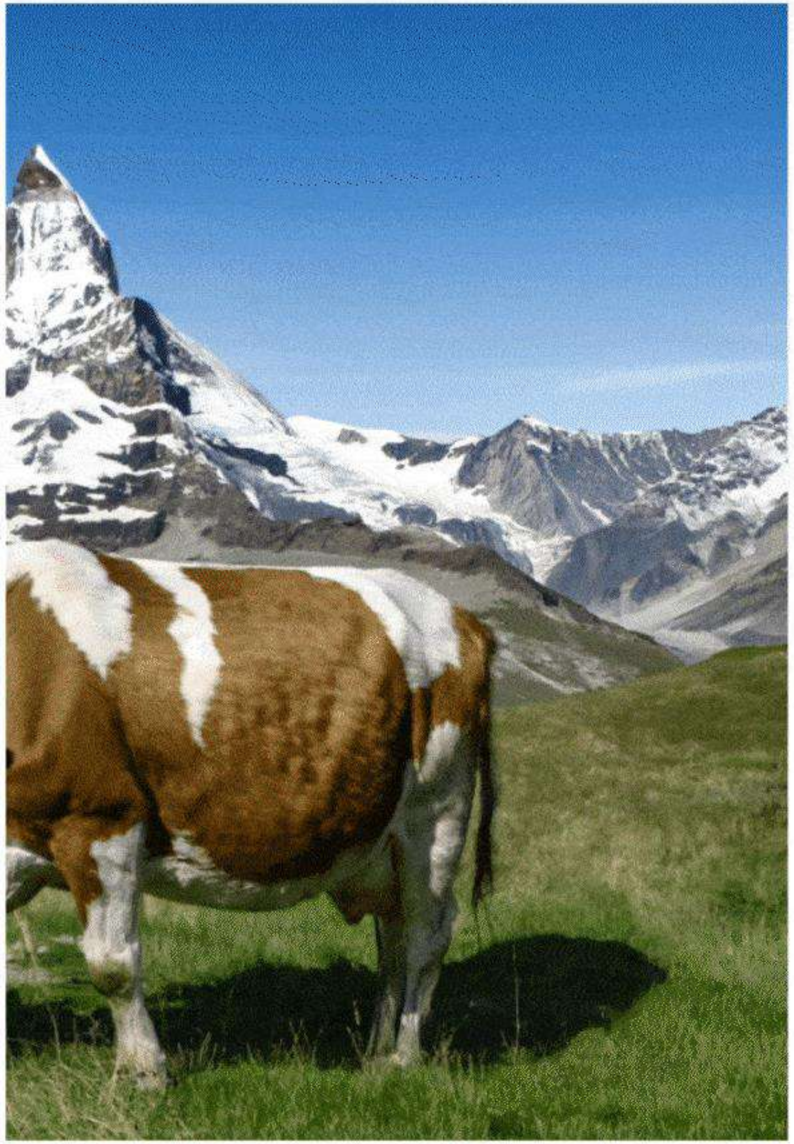
March 23, 2023

Large languag lives ranging
from browsing, to r vast societal
impacts.[1, 2, 3, 4 ie GPT family
of models.[8, 9, 1 l's limitations
(e.g., producing c sed adeptness
at providing illici nt behaviors).
Second, we give a repare GPT-4
for deployment. product- and
system-level interventions (such as monitoring and policies), and external expert engagement.
Finally, we demonstrate that while our mitigations and processes alter GPT-4's behavior and
prevent certain kinds of misuses, they are limited and remain brittle in some cases. This points
to the need for anticipatory planning and governance.[11]

"*AI systems will have even greater potential to reinforce entire ideologies, worldviews, truths and untruths, and to cement them or lock them in*".

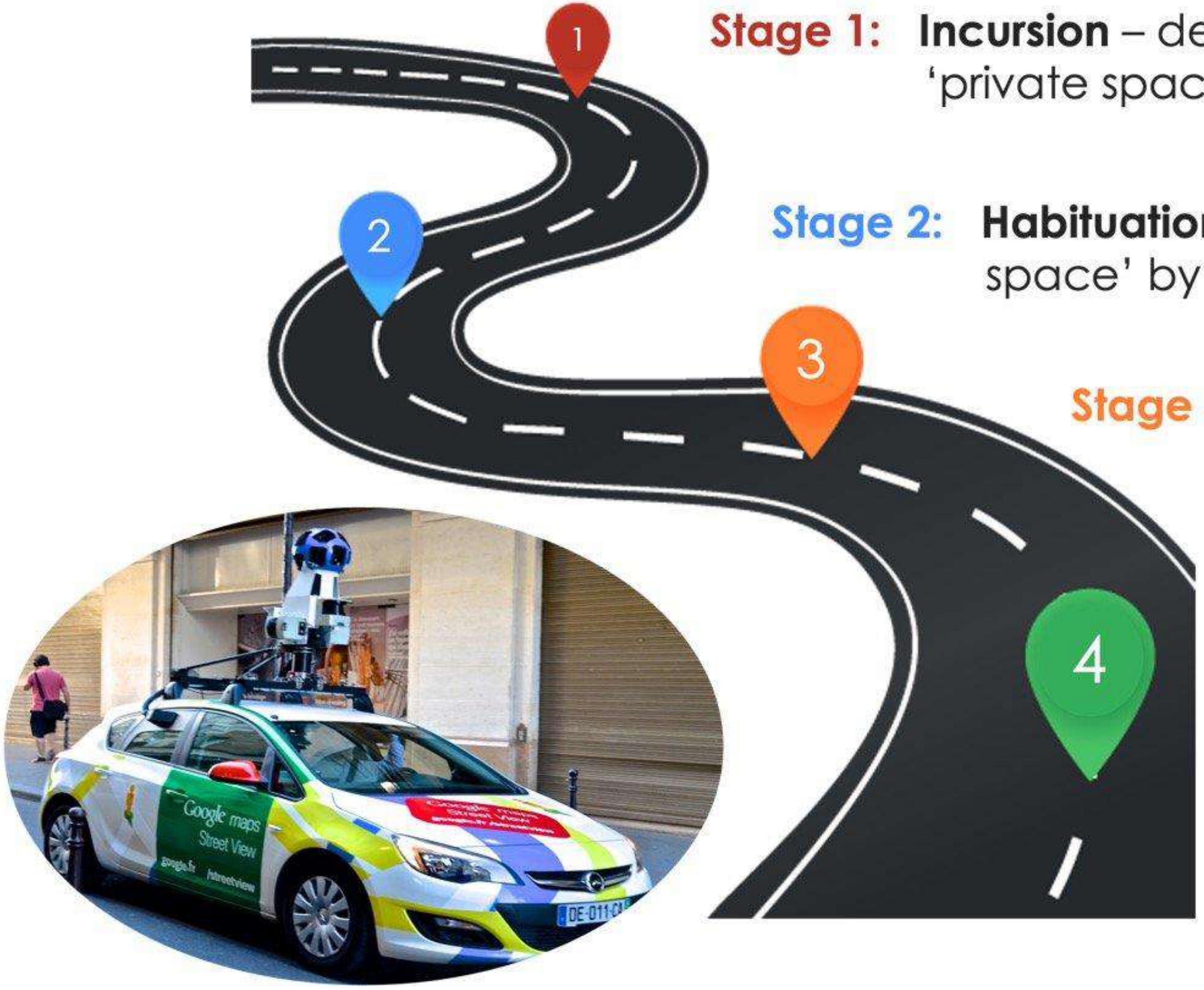


Watermarked



Non-watermarked

Dispossession Cycle



Stage 1: **Incursion** – deliberate, uninvited intrusion into a hitherto 'private space' to gather our data.

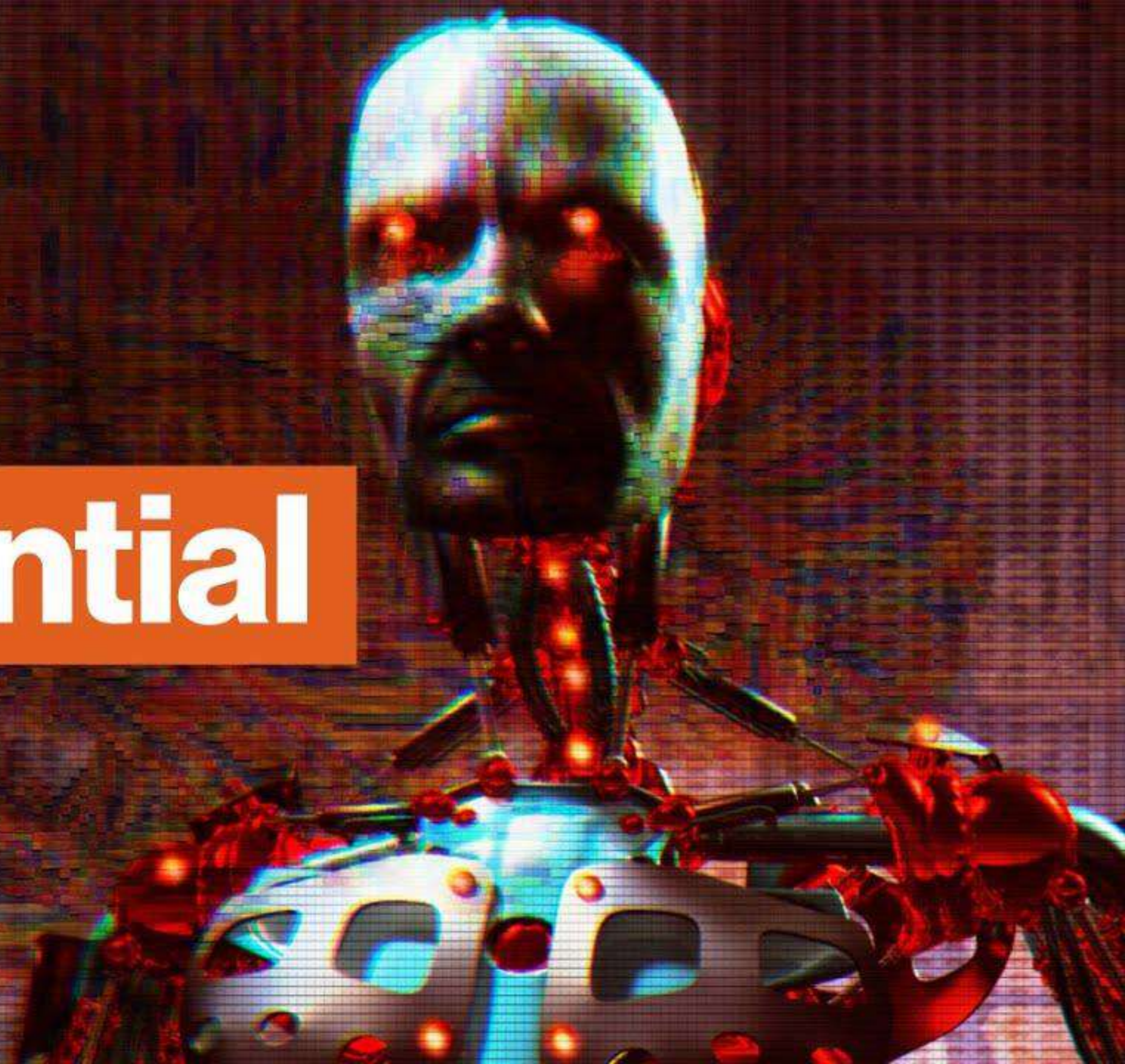
Stage 2: **Habituation** – sustained occupation of that 'private space' by the intruder for as long as possible.

Stage 3: **Adaptation** – public announcement in the face of outrage offering modified practice.

Stage 4: **Redirection** – diversion of public focus, while reverting to largely unmodified behaviour.

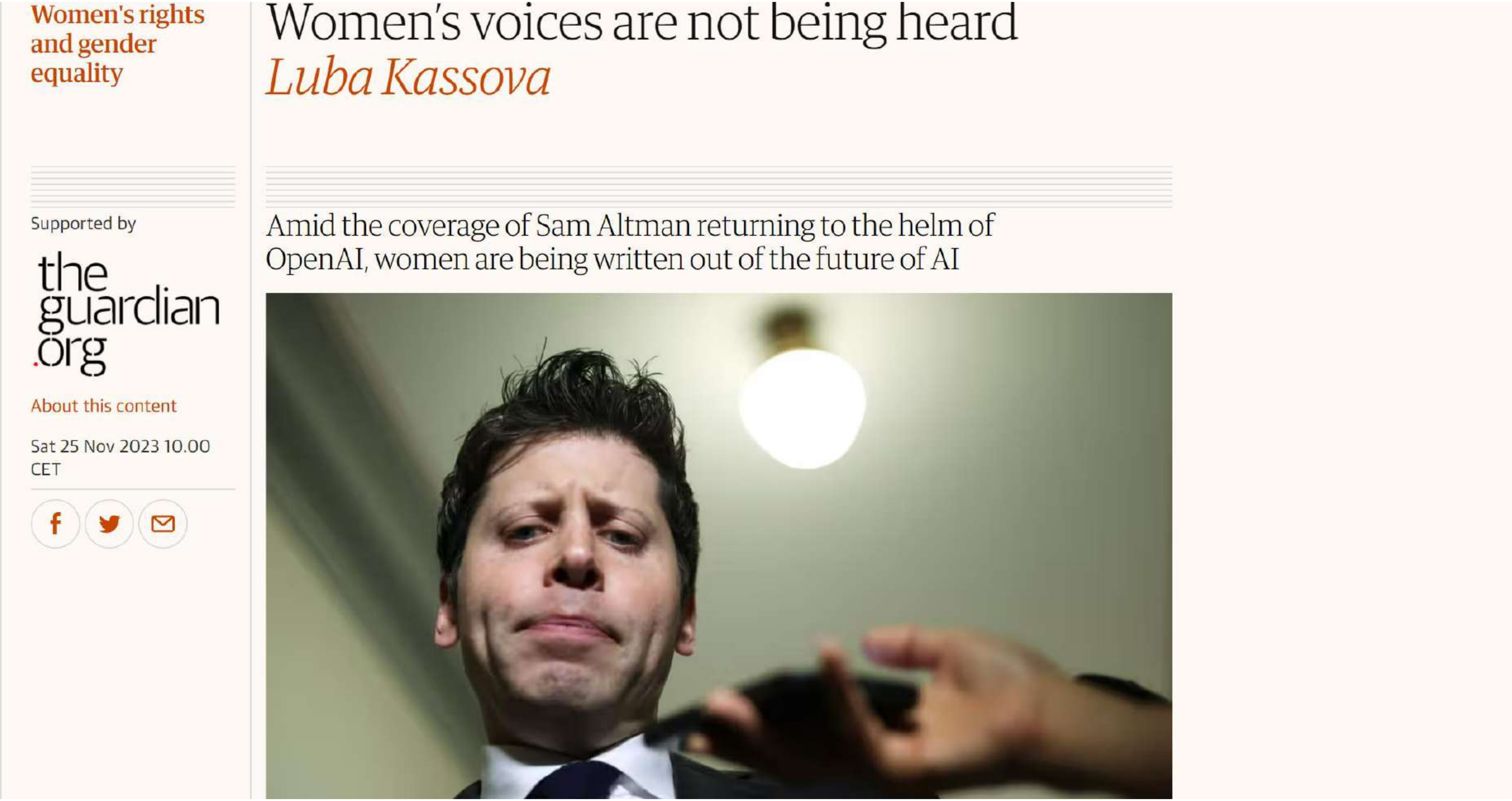
BIG THINK

Is AI an existential threat?



Statement from the listed authors of Stochastic Parrots on the "AI pause" letter di Timnit Gebru, Emily M. Bender, Angelina McMillan-Major, Margaret Mitchell:

It is indeed time to act: but the focus of our concern should not be imaginary "powerful digital minds." Instead, we should focus on the very real and very present exploitative practices of the companies claiming to build them, who are rapidly centralizing power and increasing social inequities.



Women's rights and gender equality

Women's voices are not being heard

Luba Kassova

Amid the coverage of Sam Altman returning to the helm of OpenAI, women are being written out of the future of AI

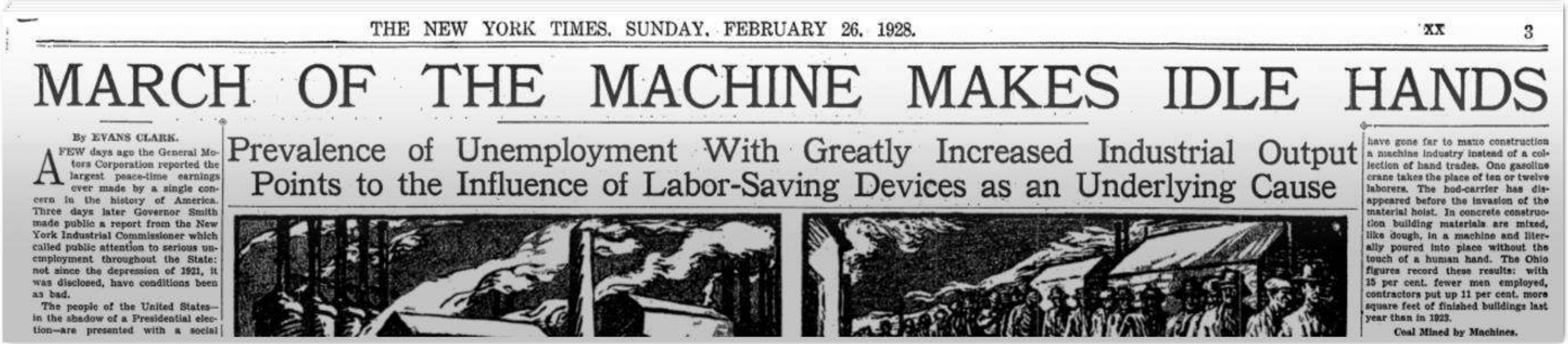
Supported by

the guardian .org

About this content

Sat 25 Nov 2023 10.00 CET

The future of work



THE NEW YORK TIMES, SUNDAY, FEBRUARY 26, 1928. XX 3

MARCH OF THE MACHINE MAKES IDLE HANDS

Prevalence of Unemployment With Greatly Increased Industrial Output Points to the Influence of Labor-Saving Devices as an Underlying Cause

By EVANS CLARK.

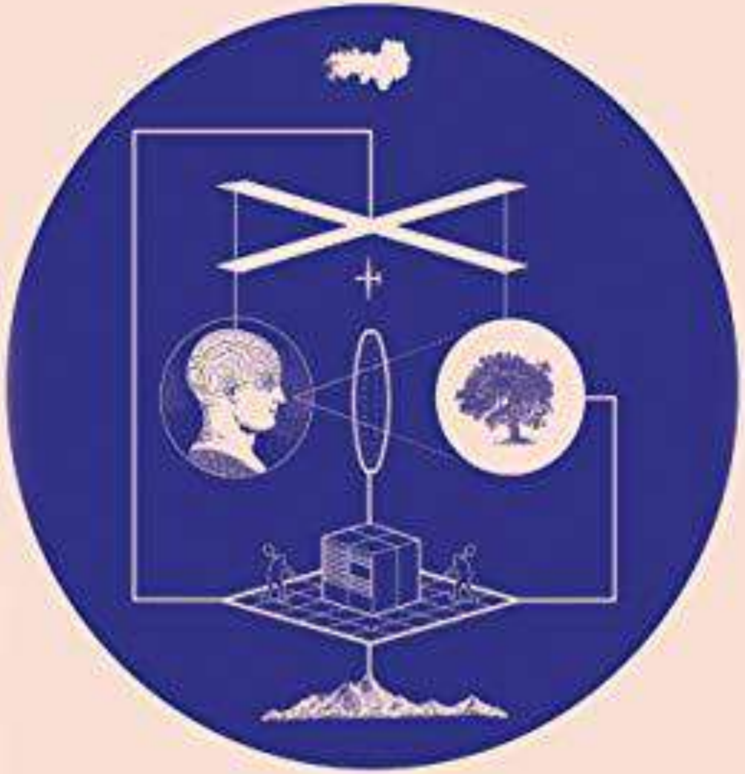
A FEW days ago the General Motors Corporation reported the largest peace-time earnings ever made by a single concern in the history of America. Three days later Governor Smith made public a report from the New York Industrial Commissioner which called public attention to serious unemployment throughout the State: not since the depression of 1921, it was disclosed, have conditions been as bad.

The people of the United States—in the shadow of a Presidential election—are presented with a social

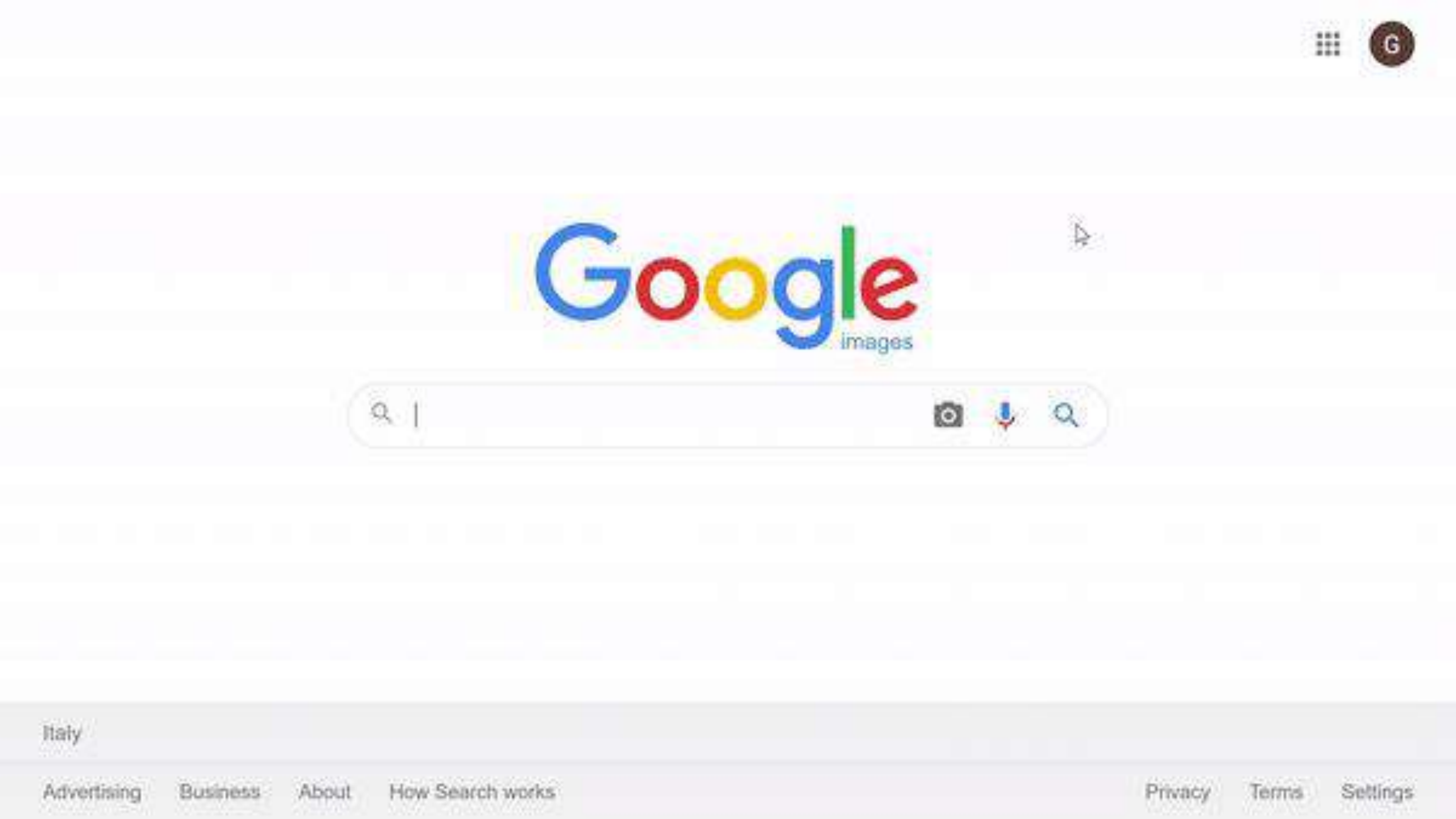
have gone far to make construction a machine industry instead of a collection of hand trades. One gasoline crane takes the place of ten or twelve laborers. The hod-carrier has disappeared before the invasion of the material hoist. In concrete construction building materials are mixed, like dough, in a machine and literally poured into place without the touch of a human hand. The Ohio figures record these results: with 35 per cent. fewer men employed, contractors put up 11 per cent. more square feet of finished buildings last year than in 1923.

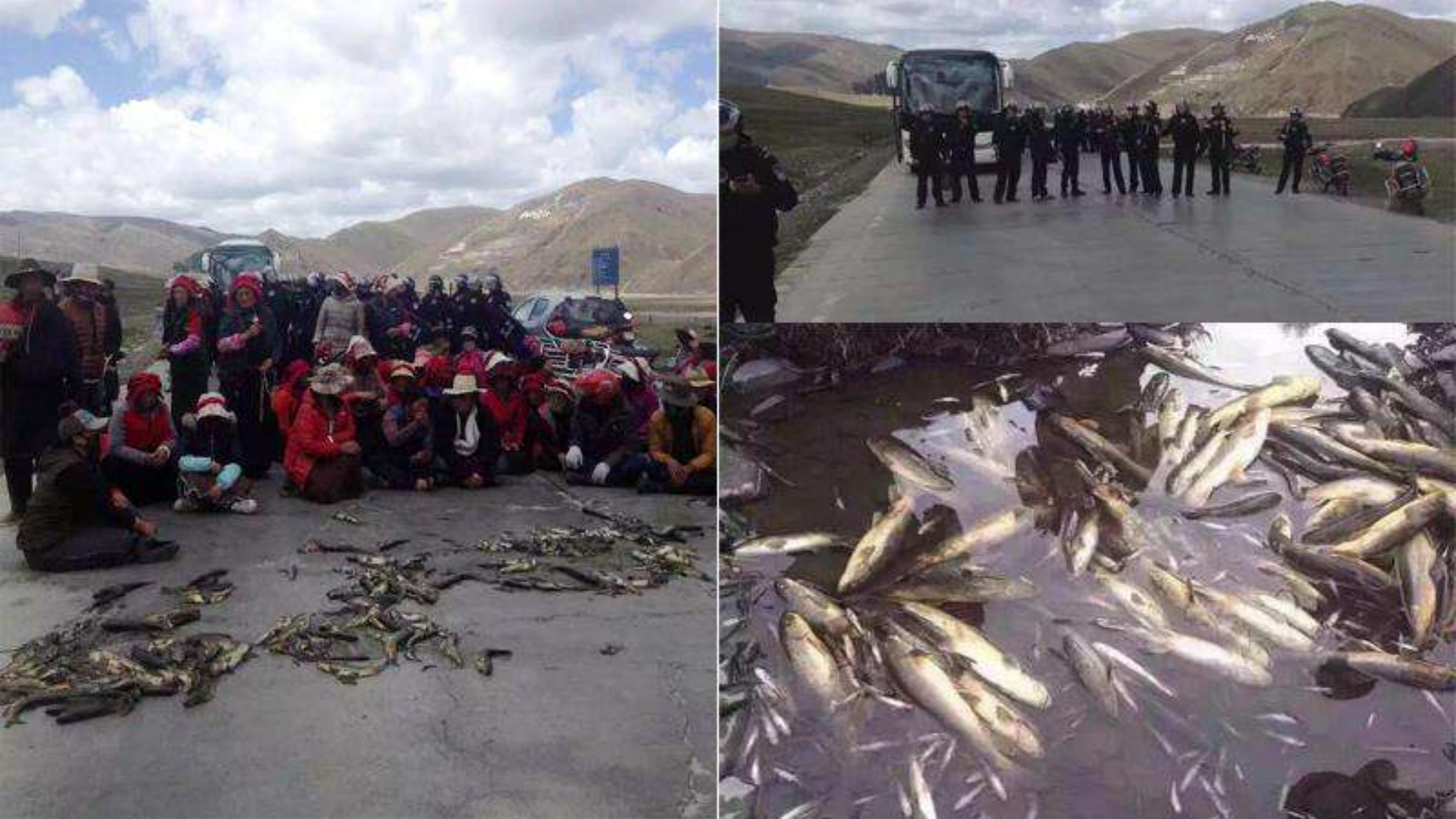
Coal Mined by Machines.

KATE CRAWFORD



ATLAS OF AI


G
Google
images
Italy
Advertising
Business
About
How Search works
Privacy
Terms
Settings



Amazon Mechanical Turk

Find an interesting task

Work

Earn Money

THE CLEANERS

TIME
CLOCK
TIME
CLOCK

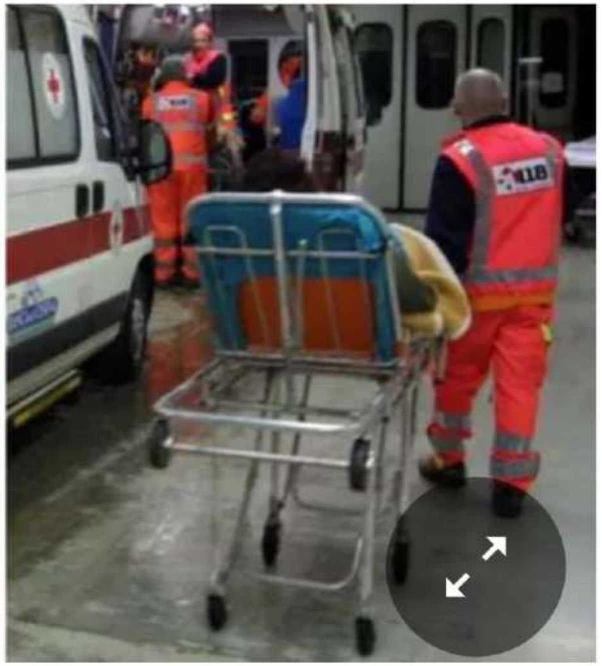
11865
11786
11750
11765

› TIKTOK SEGUI

Bimba morta per "gioco" su TikTok, oggi l'autopsia. Si indaga per istigazione al suicidio

SOCIAL

Venerdì 22 Gennaio 2021

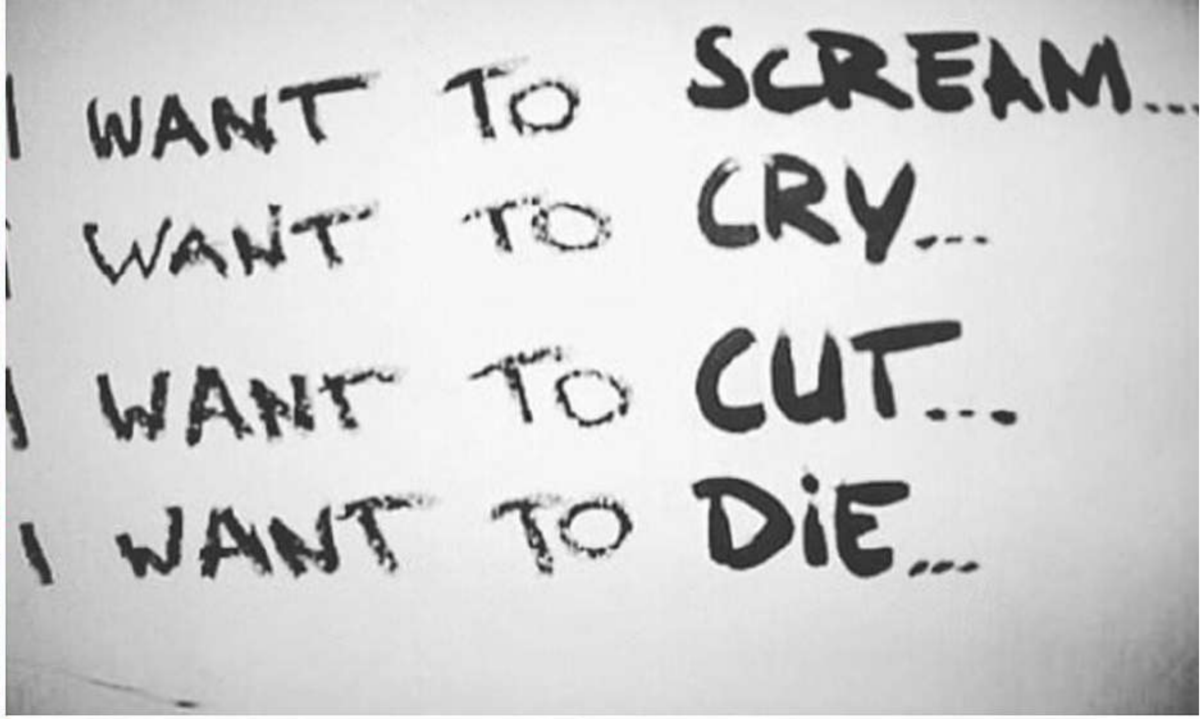


Dalla morte della bimba di 10 anni di Palermo caduta nella trappola del Blackout Challange di TikTok alla speranza di una nuova vita per più di un suo coetaneo. E' stato effettuato il prelievo degli organi di Antonella che si è strangolata durante un gioco estremo sul

MyPLAY

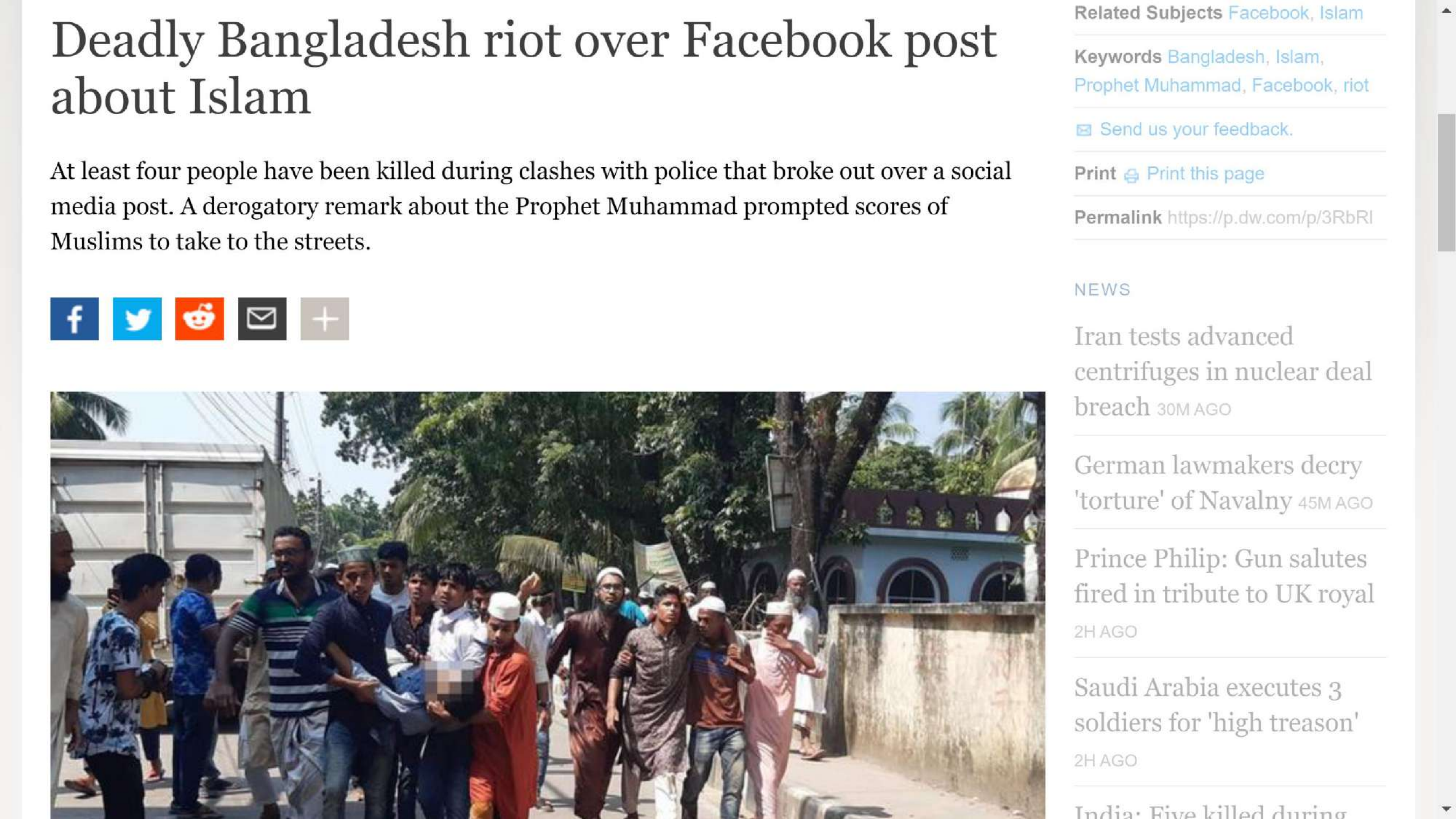
Abruzzo, in orsa e 4 cu video: anim social

Roma, cinghial



I WANT TO SCREAM...
I WANT TO CRY...
I WANT TO CUT...
I WANT TO DIE...

YOU DON'T
NEED FOOD



Deadly Bangladesh riot over Facebook post about Islam

At least four people have been killed during clashes with police that broke out over a social media post. A derogatory remark about the Prophet Muhammad prompted scores of Muslims to take to the streets.

Related Subjects Facebook, Islam

Keywords Bangladesh, Islam, Prophet Muhammad, Facebook, riot

Send us your feedback.

Print Print this page

Permalink https://p.dw.com/p/3RbRl

NEWS

Iran tests advanced centrifuges in nuclear deal breach 30M AGO

German lawmakers decry 'torture' of Navalny 45M AGO

Prince Philip: Gun salutes fired in tribute to UK royal 2H AGO

Saudi Arabia executes 3 soldiers for 'high treason' 2H AGO

India: Five killed during

Hate Speech on Facebook Is Pushing Ethiopia Dangerously Close to a Genocide

Ethnic violence set off by the assassination of a popular singer has been supercharged by hate speech and incitements shared widely on the platform.

DG By David Gilbert

September 14, 2020, 7:54pm Share Tweet Snap

MORE LIKE THIS

World News

How Did One of North Africa's Biggest Human Traffickers Escape Prison?

SALLY HAYDEN

02.23.21

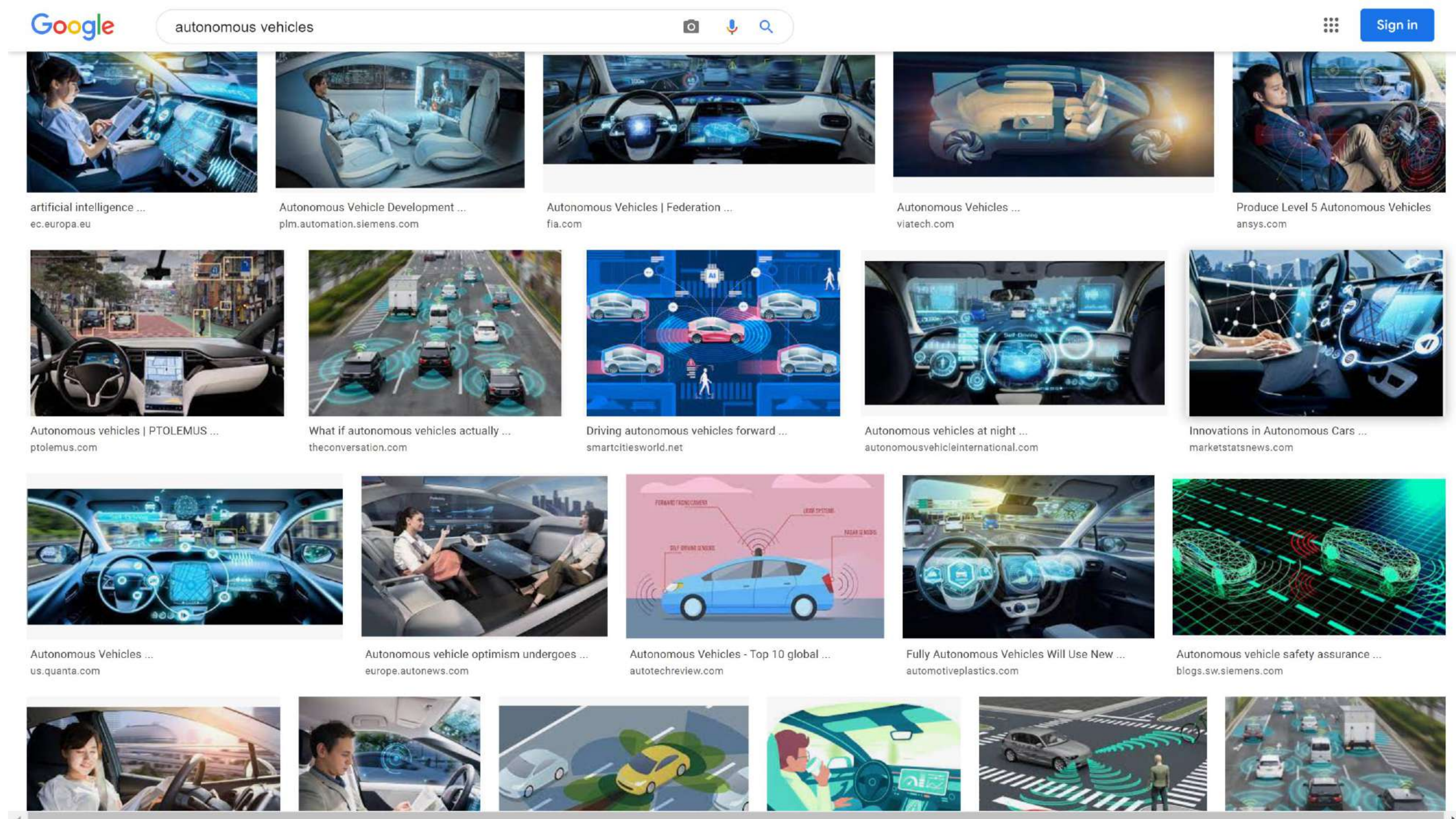
A Genocide Incited on Facebook, With Posts From Myanmar's Military

116

TRUMP
IS MY
PRESIDENT

Ethics applied to AI

- Bias and classification
- Manipulation of user
- Explainability
- Autonomy
- Privacy and surveillance
- Rhetoric
- Economics
- Society



autonomous vehicles

artificial intelligence ...
ec.europa.eu

Autonomous Vehicle Development ...
plm.automation.siemens.com

Autonomous Vehicles | Federation ...
fia.com

Autonomous Vehicles ...
viatech.com

Produce Level 5 Autonomous Vehicles
ansys.com

Autonomous vehicles | PTOLEMUS ...
ptolemus.com

What if autonomous vehicles actually ...
theconversation.com

Driving autonomous vehicles forward ...
smartcitiesworld.net

Autonomous vehicles at night ...
autonomousvehicleinternational.com

Innovations in Autonomous Cars ...
marketstatsnews.com

Autonomous Vehicles ...
us.quanta.com

Autonomous vehicle optimism undergoes ...
europe.autonews.com

Autonomous Vehicles - Top 10 global ...
autotechreview.com

Fully Autonomous Vehicles Will Use New ...
automotiveplastics.com

Autonomous vehicle safety assurance ...
blogs.sw.siemens.com

“ emergency braking maneuvers are not enabled while the vehicle is under computer control, to reduce the potential for erratic vehicle behavior “ NTSB

Special report

Apr 10th 2021 edition ›

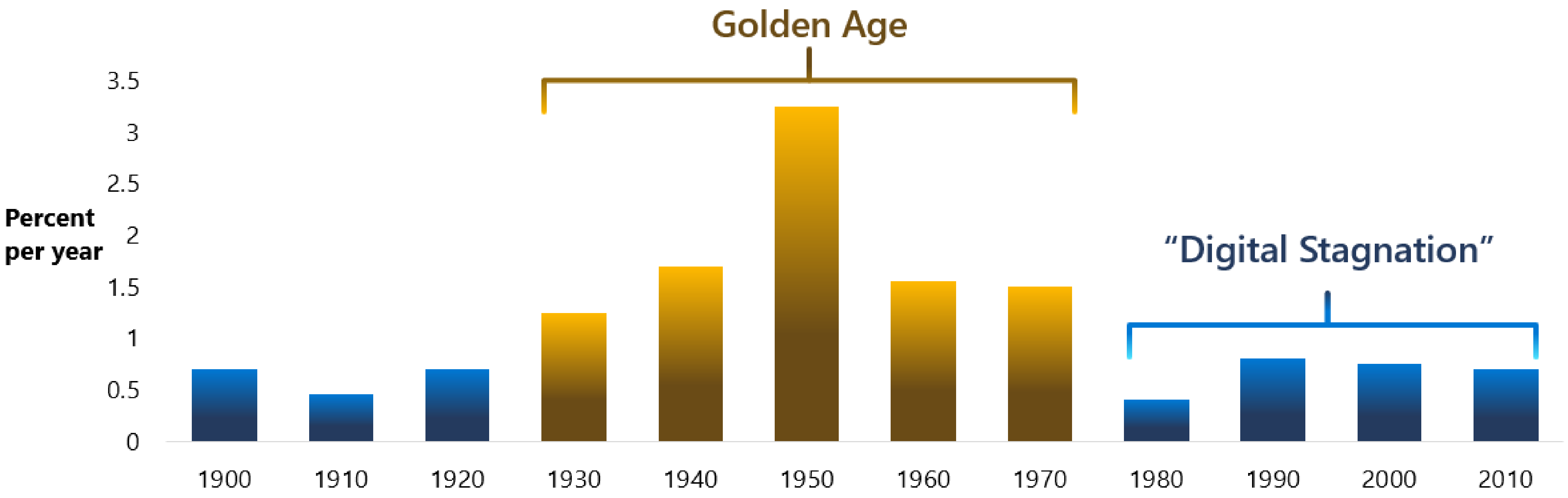
Automation

Robots threaten jobs less than fearmongers claim

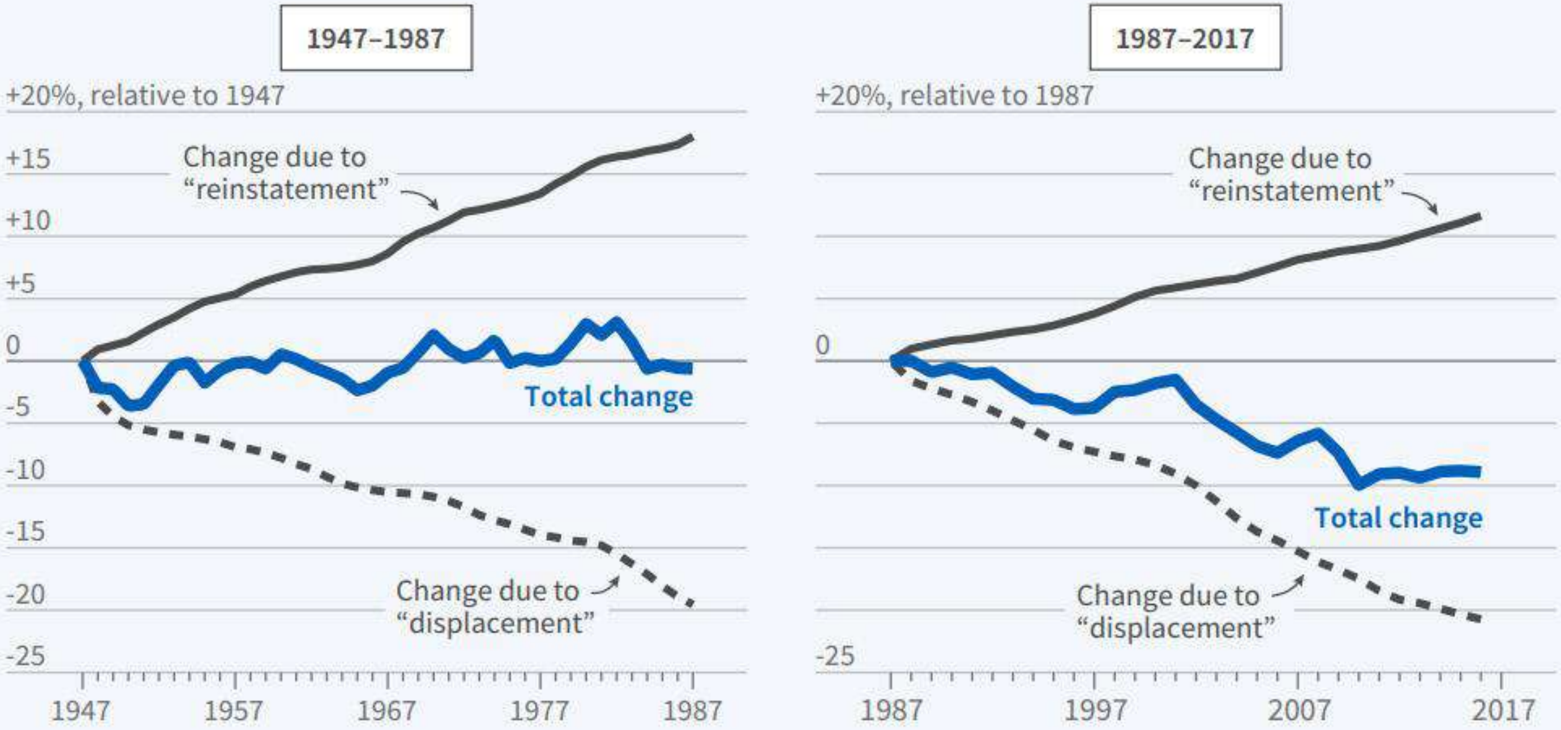
Recessions and pandemics accelerate automation. Yet warnings of a jobless future are overblown

Annual growth rate of Total Factor Productivity (TFP)

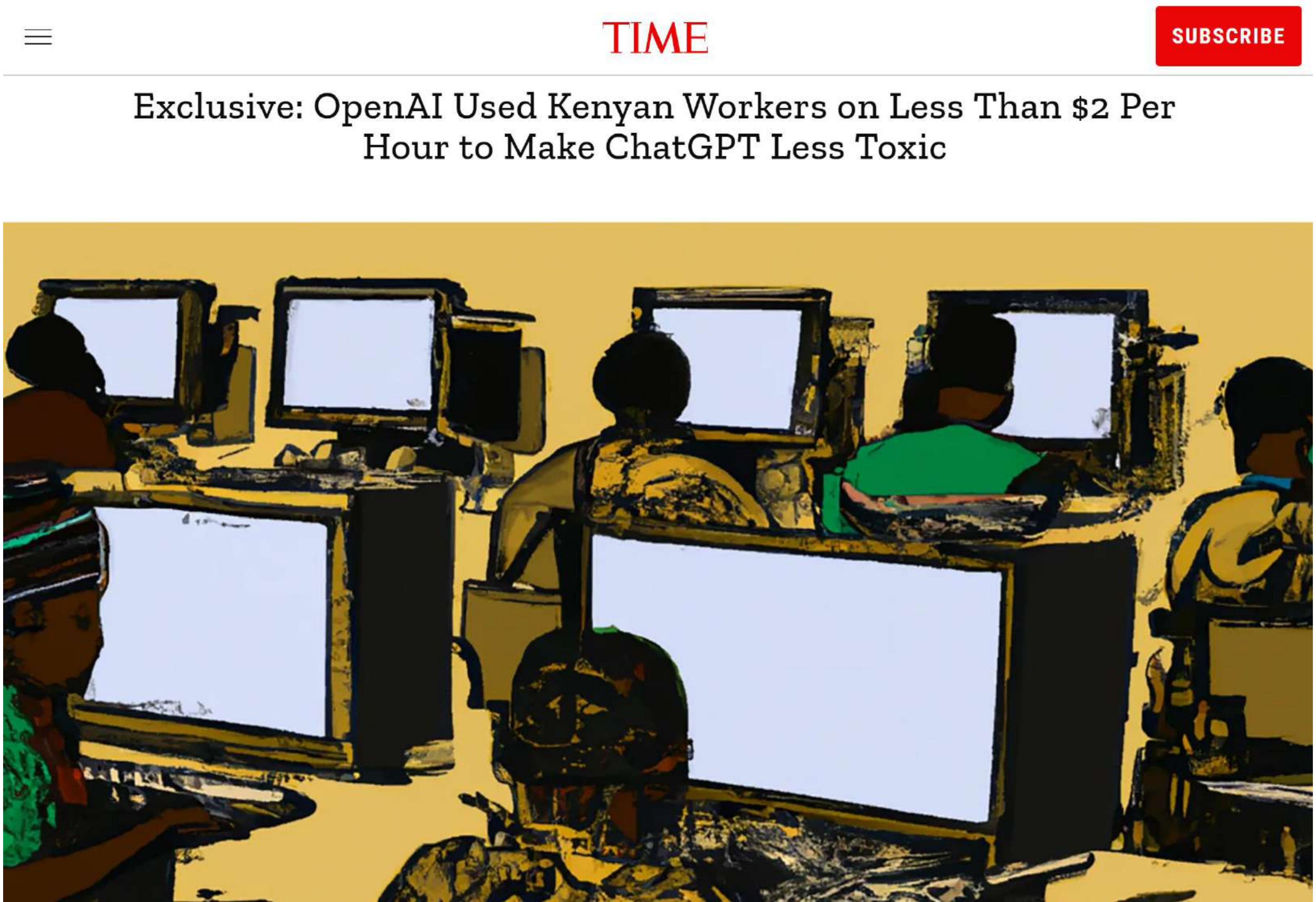
Measuring 10 years preceding years shown



Labor Share of Output and the Changing Task Content of Production



The "displacement" effect occurs when capital takes over tasks previously performed by labor.
The "reinstatement" effect occurs when technologies create new tasks in which labor has a comparative advantage.
Source: Researchers' calculations using data from the Federal Reserve Bank of St. Louis,
the Bureau of Economic Analysis, and the Bureau of Labor Statistics



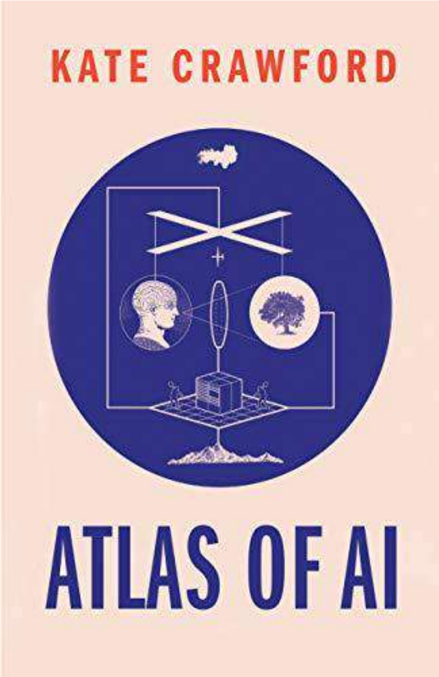
TIME

SUBSCRIBE

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

Bias or not bias

- The term "**bias**" refers to a type of **error** that can occur during this predictive process of generalization—namely, a **systematic or consistently reproduced classification error** that the system exhibits when presented with **new examples.**



Meaning of «bias»

- Outside of machine learning, “bias” has **many other meanings**.
- In law, bias refers to a **preconceived notion or opinion**, a judgment based on **prejudices**, as opposed to a decision come to from the impartial evaluation of the facts of a case.
- In psychology, Amos Tversky and Daniel Kahneman study “cognitive biases,” or the ways in which **human judgments deviate systematically from probabilistic expectations**.
- Implicit biases emphasizes the ways that **unconscious attitudes and stereotypes** “produce **behaviors** that **diverge from a person’s avowed or endorsed beliefs or principles**.”
- Here bias is **not simply a type of technical error**; it also opens onto **human beliefs**, **stereotypes**, or forms of **discrimination**. These definitional distinctions **limit the utility of “bias” as a term**, especially when used by practitioners from different disciplines.

Bias: origin

Ruha Benjamin. 2019. Race After Technology: Abolitionist Tools for the New Jim Code. Polity Press, Cambridge, UK:

“Feeding AI systems on the world’s **beauty, ugliness, and cruelty**, but expecting it to reflec **only the beauty** is a **fantasy**.”

Bias: consequences

- Consequences:
  - Quality of the results
  - Negative effect on users (e.g., discrimination from a decision system)
  - Psychological effects on users (e.g., feeling discriminated from racial assertions generated by AI systems and not knowing why)
  - Proselitism of users convinced by the products of biased AI systems (e.g., autocomplete on search proposing racist content)
  - Feedback loop reinforcing bias in the data due to data generated by biased AI systems added to the learning corpus.

Classification

- **Classification is an act of power**, be it labeling images in AI training sets, tracking people with facial recognition.
- **The practice of classification is centralizing power**: the power to decide **which differences make a difference**.

When people are categorized like objects.

Bowker and Star also underscore that **once classifications of people are constructed**, they can **stabilize a contested political category** in ways that are difficult to see.

They become **taken for granted** unless they are **actively resisted**.

We see this phenomenon in the AI field when highly **influential infrastructures and training datasets pass as purely technical**, whereas in fact **they contain political interventions** within their taxonomies: they **naturalize a particular ordering of the world** which produces **effects** that are seen **to justify their original ordering.**

Classification

- Geoffrey Bowker and Susan Leigh Star:
“classifications are powerful technologies. Embedded in working infrastructures they become relatively **invisible** without losing any of their **power**.” They can disappear “into infrastructure, into habit, into the **taken for granted**.”
- We can easily **forget** that the classifications that are casually chosen to shape a technical system can **play a dynamic role in shaping the social and material world**.

“digital epidermalization”

- IBM’s researchers go on to state an even more problematic conclusion: “**Aspects** of our **heritage**—including race, ethnicity, culture, geography—and our individual identity—age, gender and visible forms of self-expression—are **reflected in our faces**.”
- This claim **goes against decades of research** that has challenged the idea that race, gender, and identity **are biological categories** at all but are better **understood as politically, culturally, and socially constructed**.
- **Embedding identity claims** in technical systems as though they are facts observable from the face is an example of what Simone Browne calls “**digital epidermalization**,” the imposition of race on the body. Browne defines this as the **exercise of power** when the **disembodied gaze of surveillance technologies** “do the work of **alienating the subject** by producing **a ‘truth’ about the body and one’s identity** (or identities) **despite the subject’s claims**.”

The asymmetry of power

- **Technical designs** can certainly be **improved** to better account for how their systems produce skews and discriminatory results.
- But the harder questions of **why AI systems perpetuate forms of inequity** are commonly skipped over in the rush to arrive at **narrow technical solutions of statistical bias** as though that is a sufficient remedy for **deeper structural problems.**
- There has been **a general failure** to address the ways in which the instruments of knowledge in **AI reflect and serve the incentives of a wider extractive economy**.
- What remains is a persistent **asymmetry of power**, where **technical systems maintain and extend structural inequality**, regardless of the intention of the **designers**.

Political, cultural, and social choices

- To **create a training set** is to take an almost **infinitely complex and varied world** and fix it into **taxonomies** composed of **discrete** classifications of individual data points, a process that requires **inherently political, cultural, and social choices**.
- By paying attention to these **classifications**, we can glimpse the various **forms of power** that are **built** into the architectures of **AI world-building.**

Dataset: mapping the world of objects

- How the **dataset is ordered and its underlying logic** for **mapping the world of objects. ImageNet**’s structure is labyrinthine, vast, and filled with curiosities. The underlying semantic structure of ImageNet was imported from **WordNet**, a database of word classifications first developed at Princeton University’s Cognitive Science Laboratory in 1985 and funded by the U.S. Office of Naval Research.
- Its **nine top-level categories** that it drew from WordNet: plant, geological formation, natural object, sport, artifact, fungus, person, animal, and miscellaneous.

Implicit assumptions

- Object → Body → Human Body. Its subcategories include “male body,” “person,” “juvenile body,” “adult body,” and “female body.” The “**adult body**” category contains the subclasses “adult **female** body” and “adult **male** body.” There is an implicit assumption here that **only “male” and “female” bodies are recognized as “natural.”**
- Address the deeper harm of **allocating people into gender or race categories without their input or consent.** This practice has a long history. **Administrative systems** for centuries have sought to make humans legible by **applying fixed labels and definite properties**.
- The work of **essentializing and ordering** on the basis of biology or culture has long been used to **justify forms of violence and oppression**.

Private dataset

- The classification schemes used in **companies** like Facebook are much harder to **investigate and criticize**: **proprietary** systems offer few ways for outsiders **to probe or audit** how images are ordered or interpreted.

Irresolvable questions

- **Images**—like all forms of data—are laden with all sorts of **potential meanings, irresolvable questions, and contradictions**.
- In trying to **resolve these ambiguities**, ImageNet’s labels **compress** and **simplify complexity**.

How should AI systems make representations of the social?

Defining categories and ideas of normalcy creates an outside: forms of abnormality, difference, and otherness.

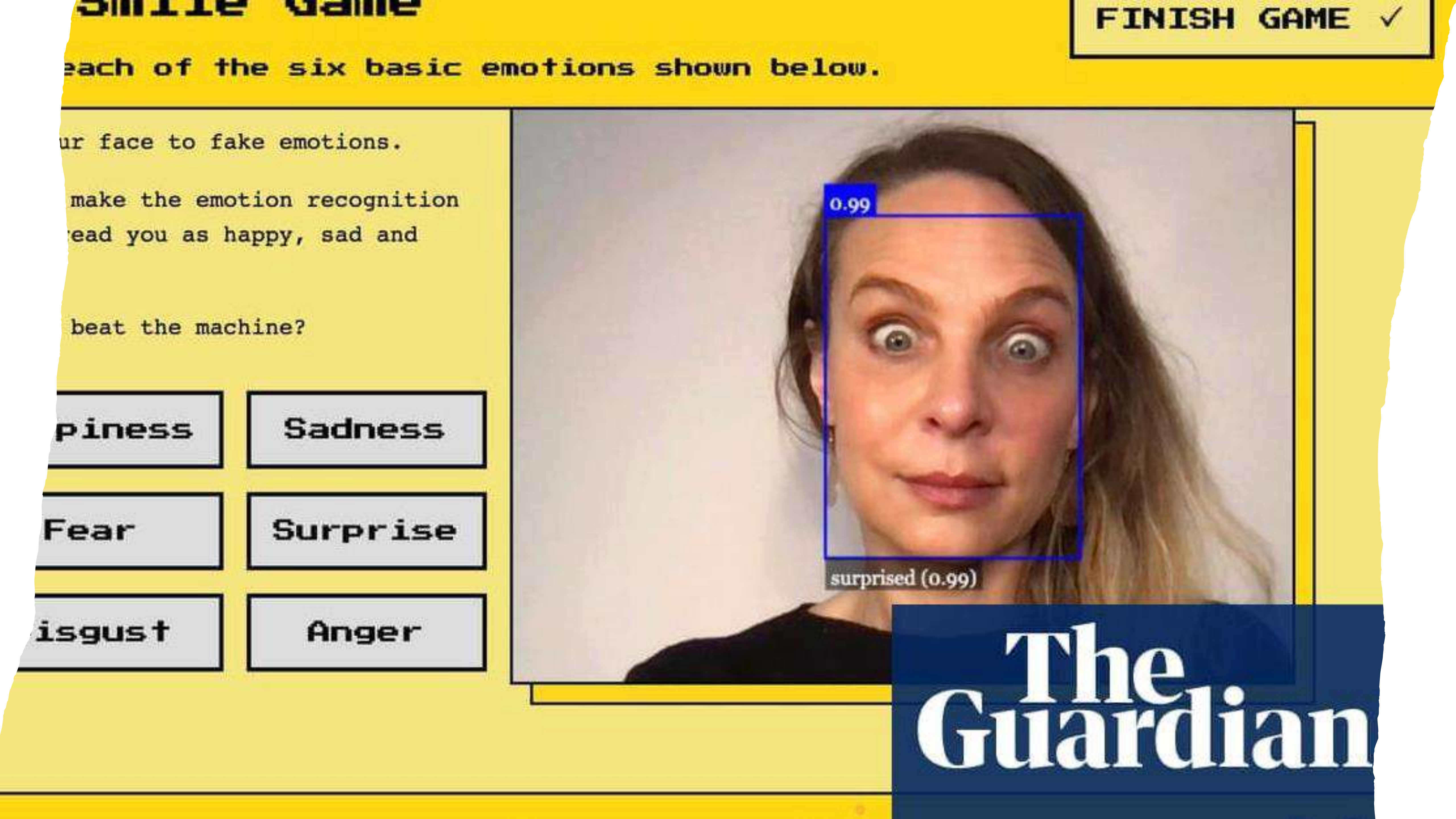
Technical systems are making **political and normative interventions** when they give names to something as **dynamic and relational as personal identity**, and they commonly do so using a **reductive set of possibilities** of **what it is to be human.**

That restricts the range of **how people are understood** and can **represent themselves**, and it **narrows** the horizon of recognizable **identities**.

As Ian Hacking observes, classifying people is an **imperial imperative**: subjects were classified by empires when they were **conquered**, and then they were ordered into "**a kind of people**" by **institutions and experts.**

Who gets to choose?

- **AI** systems to **produce new classifications** is a powerful moment of decision making: but **who gets to choose and on what basis**?
- The problem for computer science is that **justice** in AI systems will never be something that can be coded or computed. It requires a shift to **assessing systems beyond optimization metrics and statistical parity** and an understanding of where the frameworks of mathematics and engineering are causing the problems. This also means understanding how AI systems interact with **data, workers, the environment,** and the **individuals whose lives will be affected** by its use and **deciding where AI should not be used.**



Smile Game

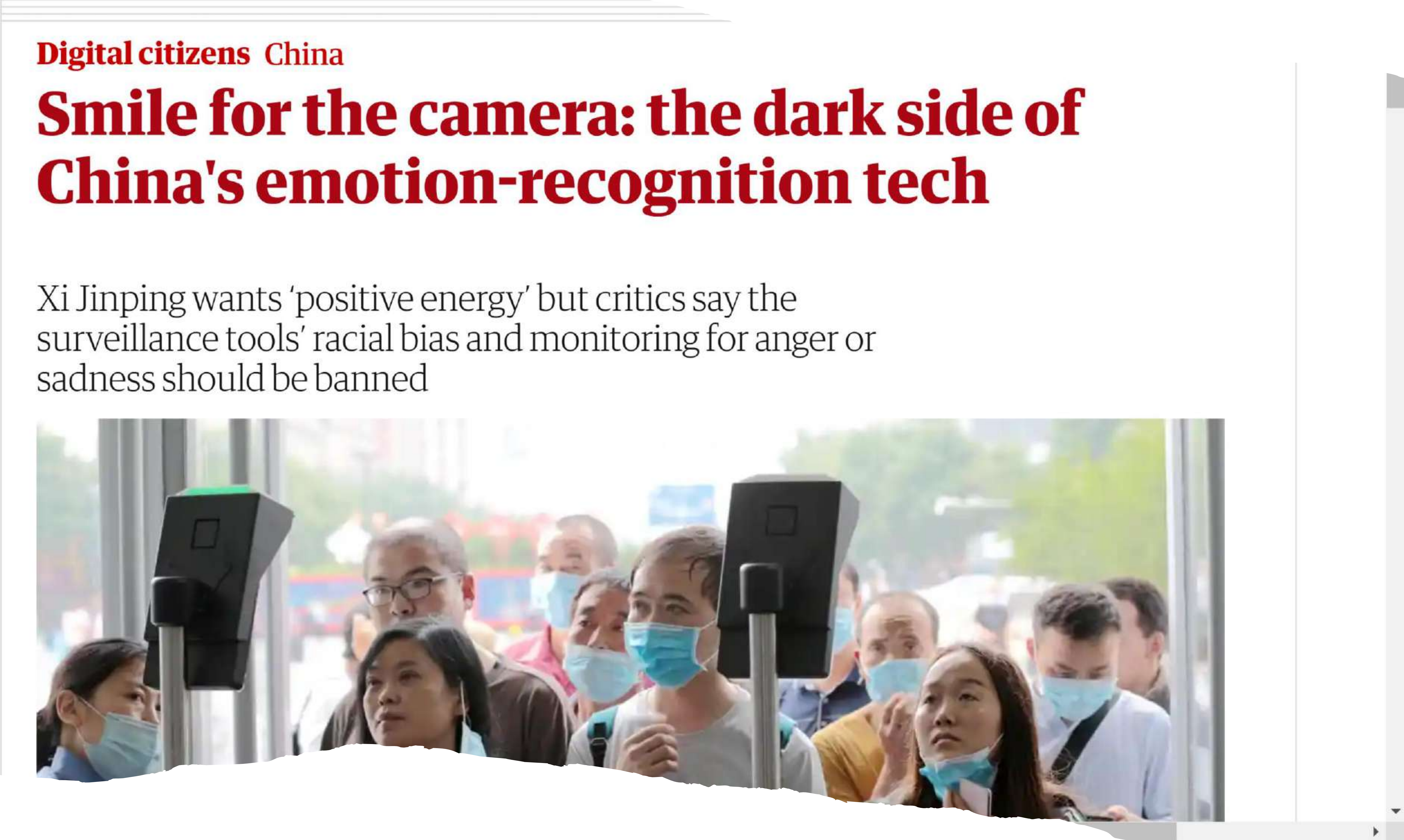
each of the six basic emotions shown below.

ur face to fake emotions.

make the emotion recognition read you as happy, sad and

beat the machine?

The Guardian



Digital citizens China

Smile for the camera: the dark side of China's emotion-recognition tech

Xi Jinping wants ‘positive energy’ but critics say the surveillance tools’ racial bias and monitoring for anger or sadness should be banned

REVISED EDITION

"You'll never look at other people in quite the same way again. *Emotions Revealed* is a tour de force."
—**Malcolm Gladwell, author of *Blink***

emotions revealed

RECOGNIZING FACES AND FEELINGS TO IMPROVE COMMUNICATION AND EMOTIONAL LIFE

Paul Ekman

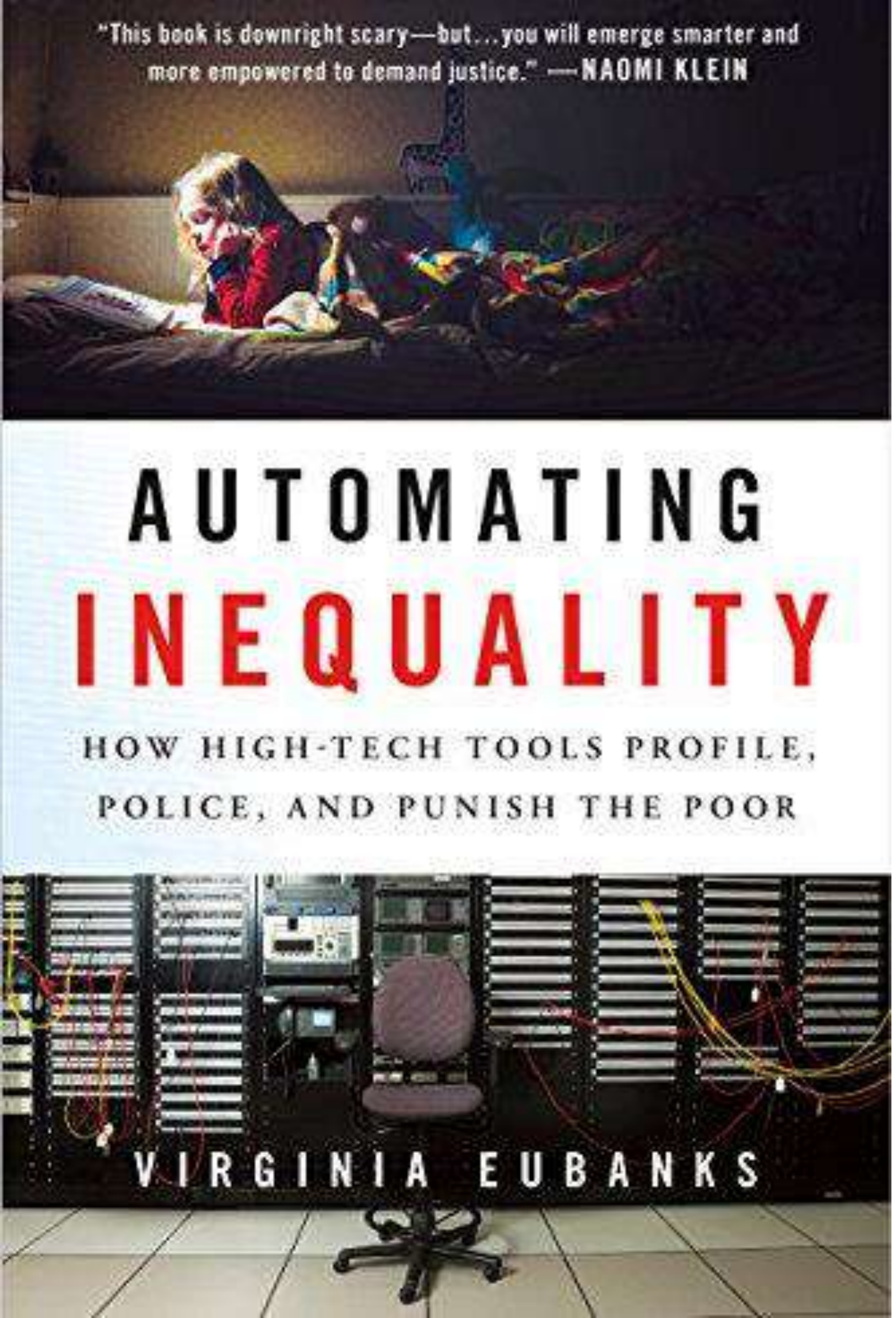
WITH A NEW CHAPTER ON EMOTIONS AND LYING

CODED BIAS

OFFICIAL SELECTION 2020

sundance

film festival

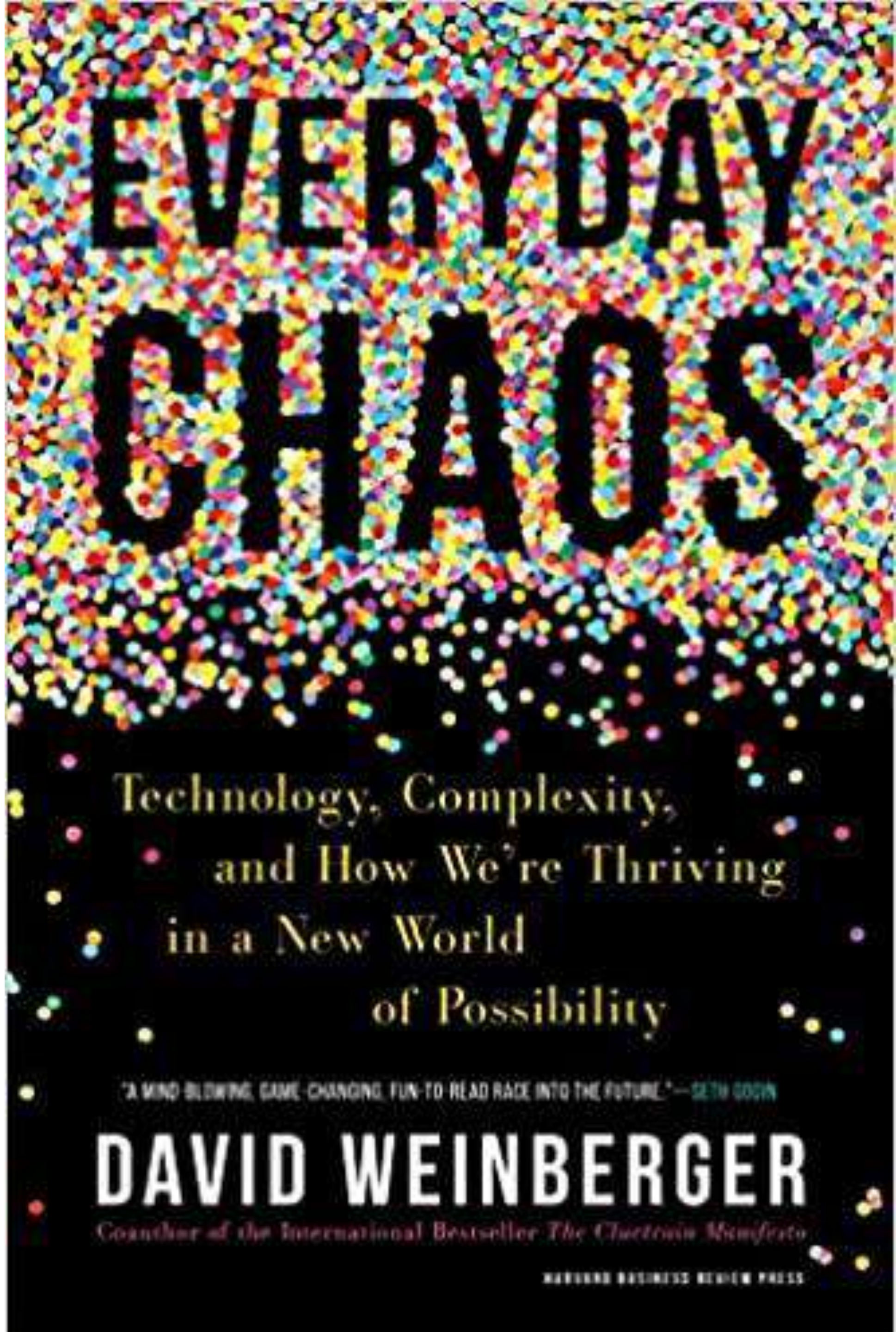


"This book is downright scary—but…you will emerge smarter and more empowered to demand justice." —NAOMI KLEIN

AUTOMATING INEQUALITY

HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR

VIRGINIA EUBANKS



EVERYDAY CHAOS

Technology, Complexity,
and How We're Thriving
in a New World
of Possibility

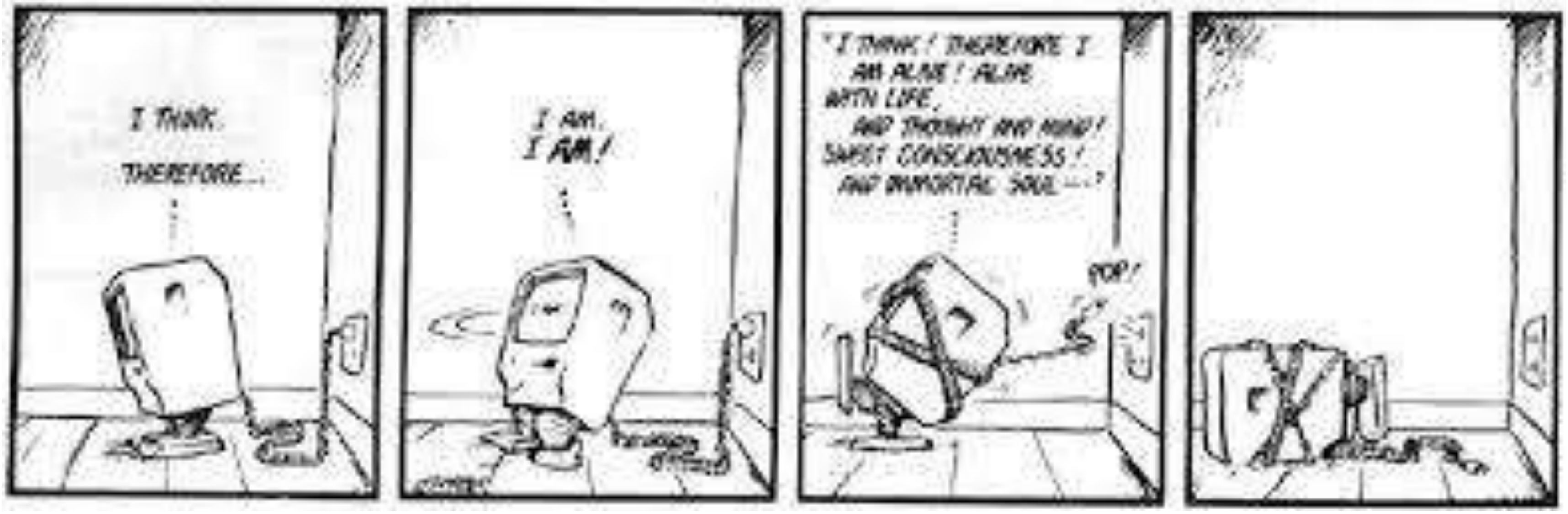
"A MIND-BLOWING, GAME-CHANGING, FUN-TO-READ RACE INTO THE FUTURE." —SETH GODIN

DAVID WEINBERGER

Coauthor of the International Bestseller *The Cluetrain Manifesto*

HARVARD BUSINESS REVIEW PRESS

Dr. Halley.


I THINK.
THEREFORE...
I AM.
I AM!
"I THINK! THEREFORE I AM ALIVE! ALIVE WITH LIFE, AND THOUGHT AND MIND! SWEET CONSCIOUSNESS! AND IMMORTAL SOUL...?
POP!

Explainability

- GDPR Recital 71:
- The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention.
- In any case, such processing should be subject to **suitable safeguards**, which should include specific information to the data subject and **the right** to obtain human intervention, to express his or her point of view**, to obtain an explanation of the decision reached after such assessment** and <u>**to challenge the decision.**</u>

Problems

- What is an explanation?
- Can humans give explanations?
- Tacit knowledge (Michael Polanyi)
- How to explore dependeces among hundreds of trillions of nodes in a neural network?
- Could you trust the machine explanation?
- Are human conceptual models only simplifications of a complex reality?

Registry of power

- **AI is neither artificial nor intelligent**. Rather, artificial intelligence is both **embodied and material**, made from natural resources, fuel, human labor, infrastructures, logistics, histories, and classifications.
- AI systems **are not autonomous**, rational, or able to discern anything without extensive, computationally intensive **training with large datasets** or predefined rules and rewards.
- In fact, **artificial intelligence** as we know it **depends entirely on a much wider set of political and social structures**.
- And due to the **capital required to build AI** at scale and the **ways** of seeing that it optimizes AI systems are ultimately **designed to serve existing dominant interests.**
- AI is a **registry of power**. How artificial intelligence is made, in the widest sense, and the economic, political, cultural, and historical forces that shape it.
- Once we connect **AI within** these **broader structures and social systems**, we can **escape** the notion that **artificial intelligence is a purely technical domain.**

AI: the massive industrial formation

- **AI** systems both **reflect and produce social relations and understandings of the world**.
- It’s worth noting that the term “artificial intelligence” can create **discomfort** in the computer science community. The phrase has moved in and out of fashion over the decades and is used more in marketing than by researchers. “**Machine learning**” is more commonly used in the technical literature.
- Yet the nomenclature of AI is often embraced during **funding** application season, when venture capitalists come bearing checkbooks, or when researchers are seeking **press** attention for a new scientific result. As a result, the term is both used and rejected in ways that keep its **meaning in flux**.
- For my purposes, I use **AI** to talk about **the massive industrial formation that includes politics, labor, culture, and capital.**

Neither artificial nor intelligent

- AI is **neither artificial nor intelligent**. Rather, artificial intelligence is both **embodied** and **material**, made from natural resources, fuel, human **labor**, infrastructures, logistics, histories, and classifications. AI systems are **not autonomous,** rational, or able to discern anything without extensive, computationally intensive **training** with large datasets or predefined rules and rewards.

Neither artificial nor intelligent

- Artificial intelligence as we know it **depends entirely on a much wider set of political and social structures**. And due to the **capital** required to build AI at **scale** and the ways of seeing that it optimizes **AI systems are ultimately designed to serve existing dominant interests**. In this sense, artificial intelligence is a **registry of power**.
- Once we connect AI within these broader structures and social systems, we can escape the notion that artificial intelligence is a **purely technical domain**. At a fundamental level, AI is technical and **social practices, institutions and infrastructures, politics and culture.**

The asymmetry of power

- **Technical designs** can certainly be **improved** to better account for how their systems produce skews and discriminatory results.
- But the harder questions of **why AI systems perpetuate forms of inequity** are commonly skipped over in the rush to arrive at **narrow technical solutions of statistical bias** as though that is a sufficient remedy for **deeper structural problems.**

The asymmetry of power

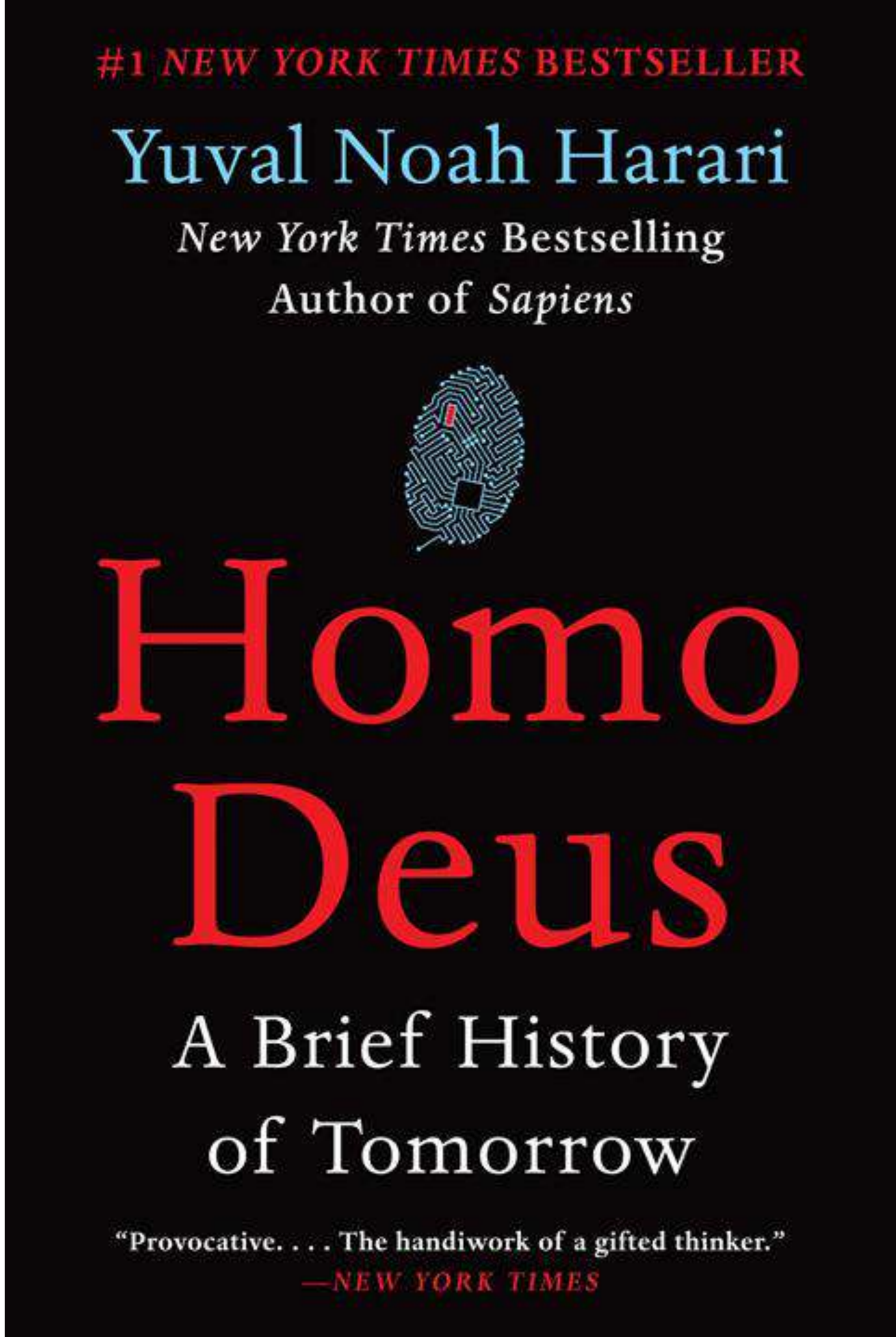
- There has been **a general failure** to address the ways in which the instruments of knowledge in **AI reflect and serve the incentives of a wider extractive economy**.
- What remains is a persistent **asymmetry of power**, where **technical systems maintain and extend structural inequality**, regardless of the intention of the **designers**.

Should we not seek to democratize it?

- If AI currently **serves the existing structures of power**, an obvious **question** might be: **Should we not seek to democratize it?** Could there not be an AI for the people that is **reoriented** toward **justice** and **equality** rather than industrial extraction and discrimination?
- This may seem **appealing**, but the **infrastructures** and forms of **power** that **enable and are enabled by AI** skew strongly toward the **centralization of control**. To suggest that we democratize AI to reduce asymmetries of power is a little like arguing for democratizing **weapons manufacturing** in the service of peace

BIG THINK

Is AI an existential threat?



#1 *NEW YORK TIMES* BESTSELLER

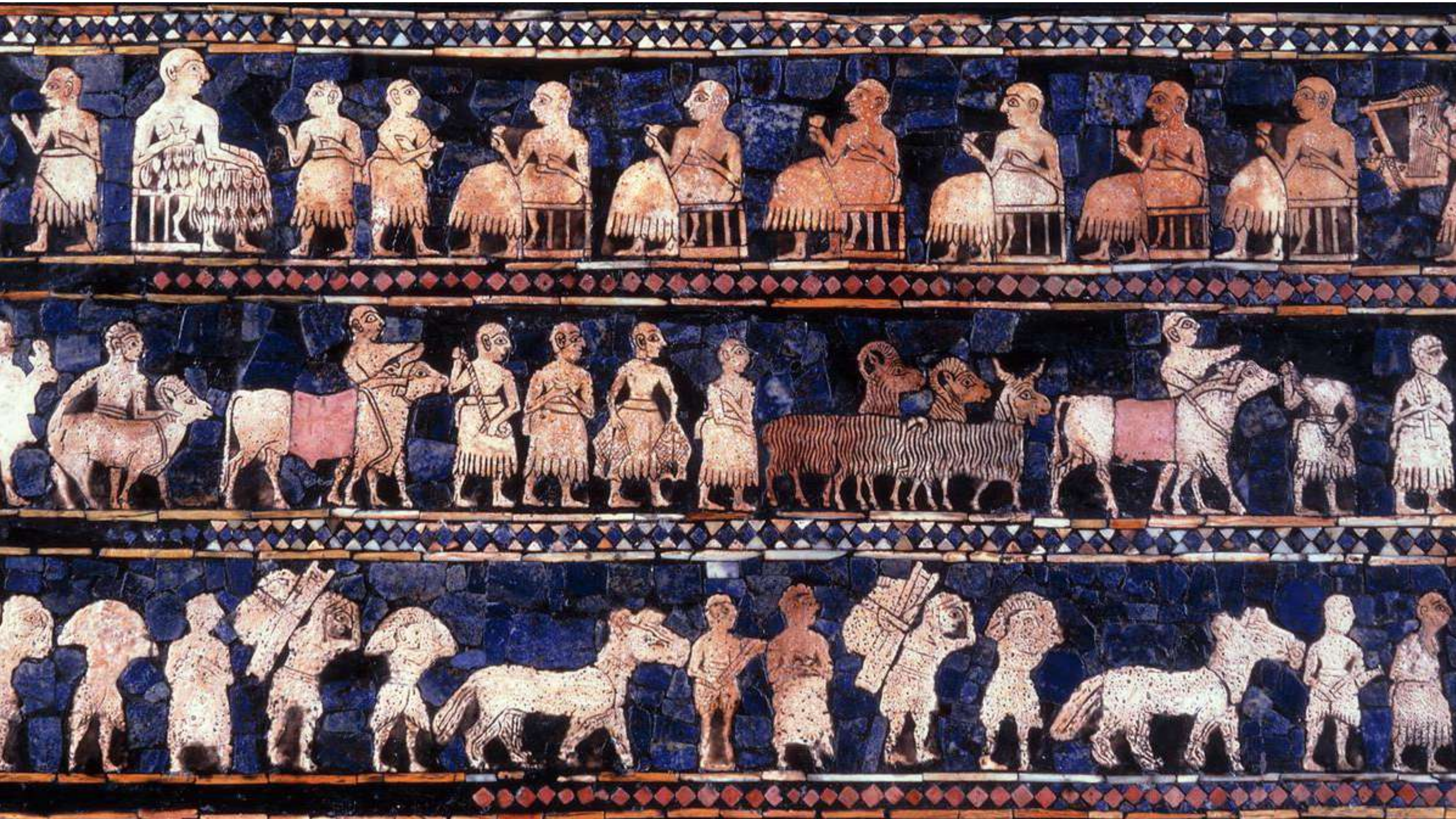
Yuval Noah Harari

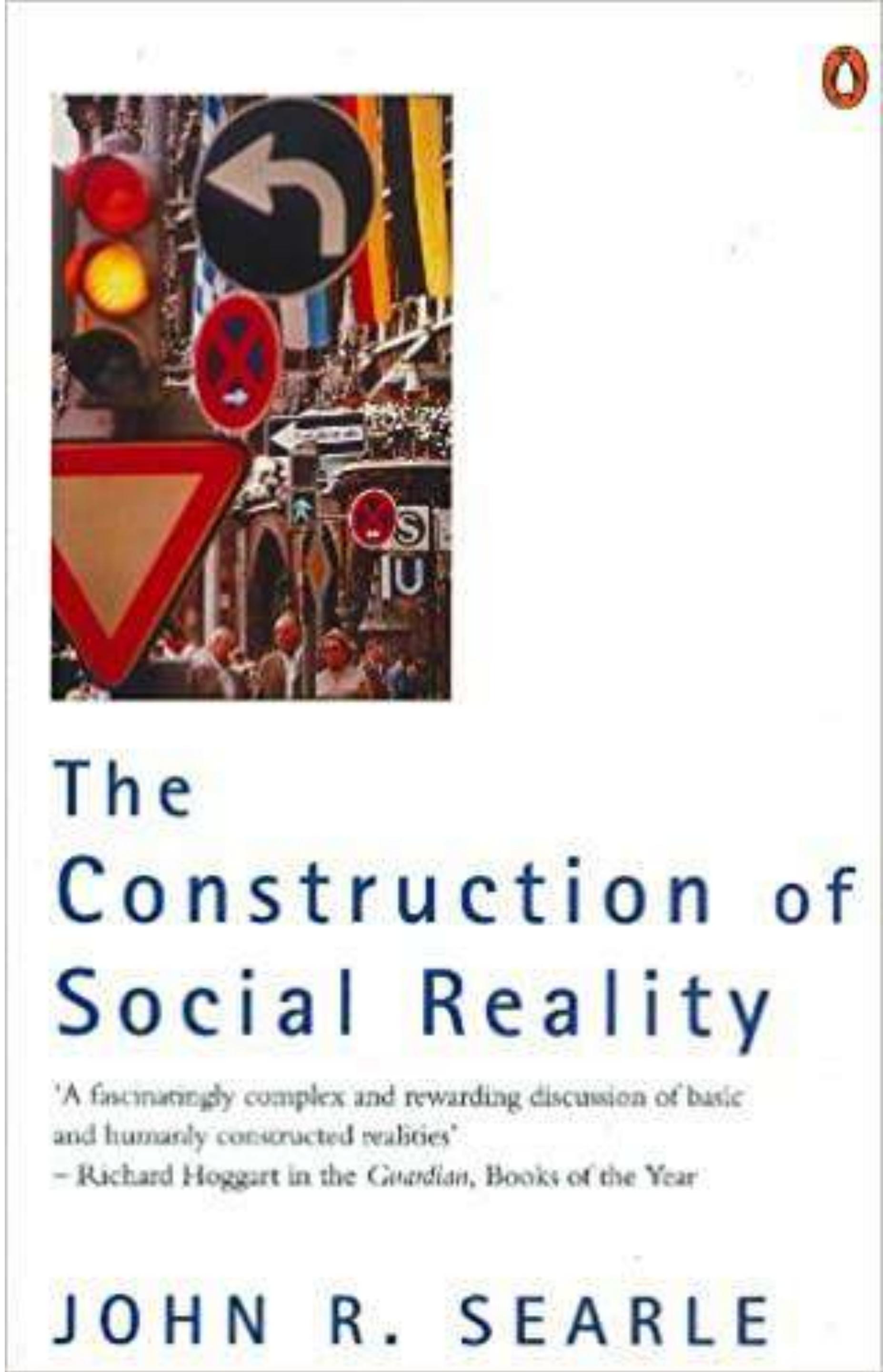
New York Times Bestselling Author of *Sapiens*

Homo Deus

A Brief History of Tomorrow

"Provocative. . . . The handiwork of a gifted thinker."
—*NEW YORK TIMES*

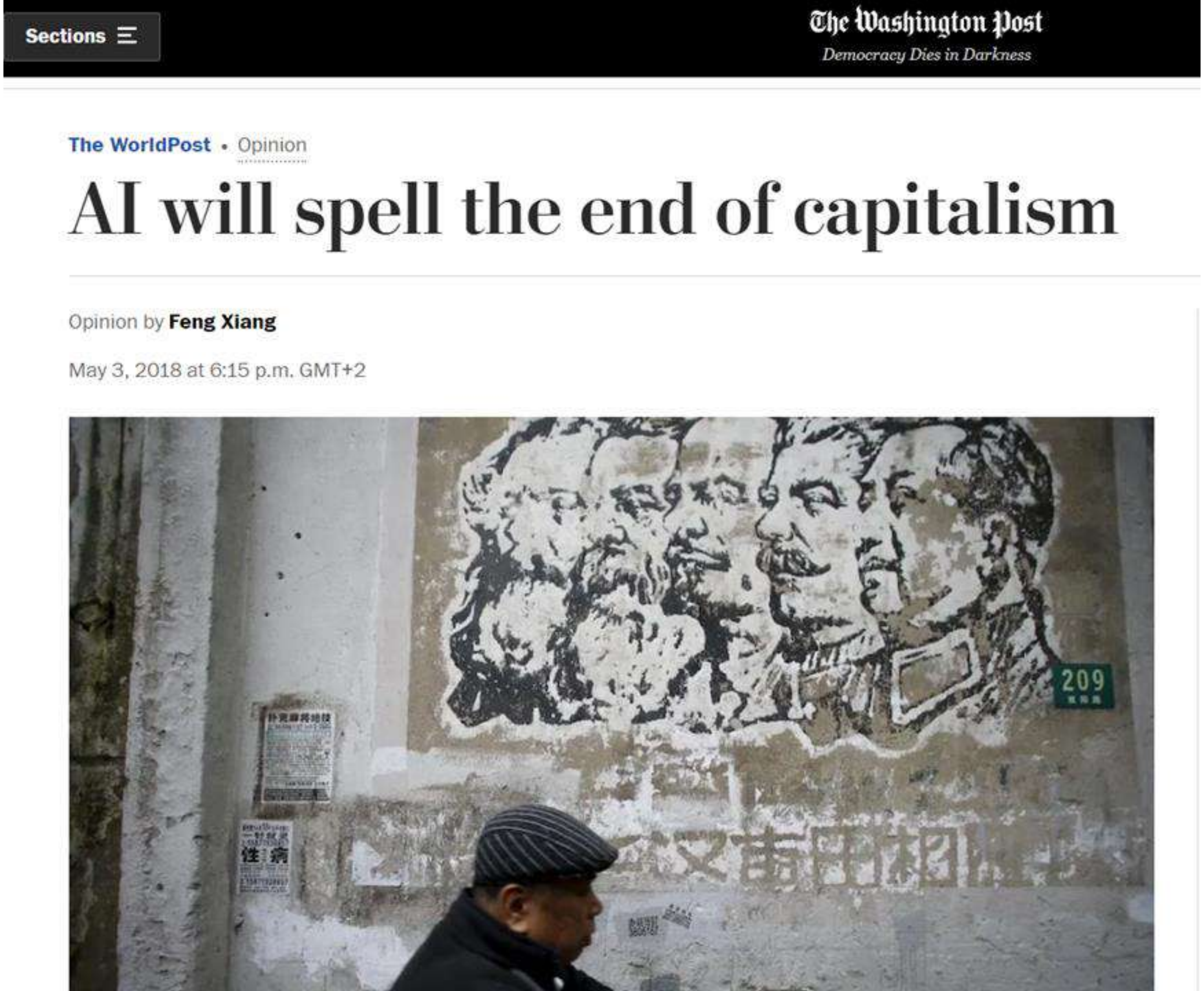




The Construction of Social Reality

'A fascinatingly complex and rewarding discussion of basic and humanly constructed realities'
– Richard Hoggart in the *Guardian*, Books of the Year

JOHN R. SEARLE



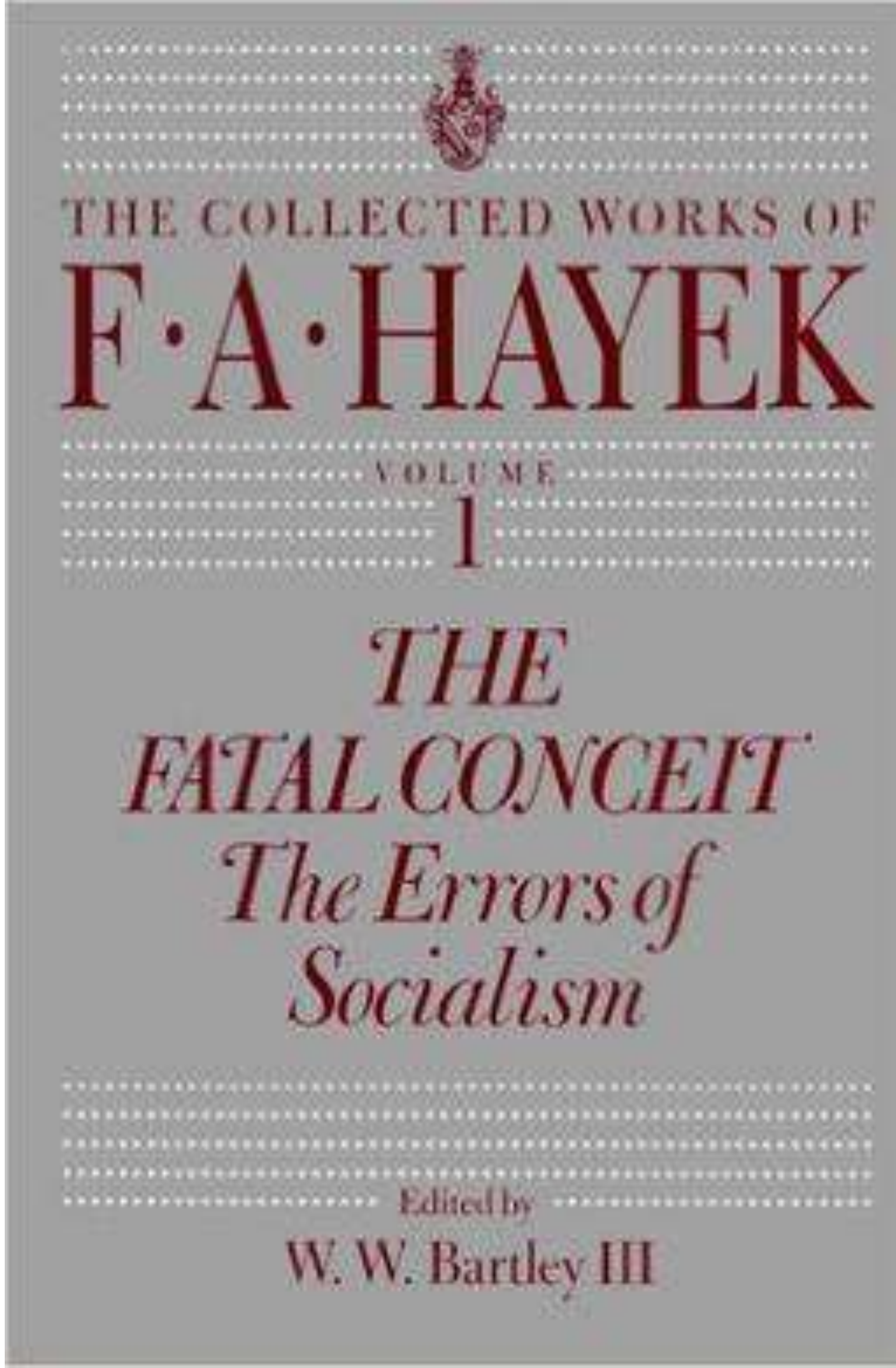
The Washington Post
Democracy Dies in Darkness

The WorldPost • Opinion

AI will spell the end of capitalism

Opinion by **Feng Xiang**

May 3, 2018 at 6:15 p.m. GMT+2



THE COLLECTED WORKS OF

F·A·HAYEK

VOLUME

1

THE

FATAL CONCEIT

The Errors of Socialism

Edited by

W. W. Bartley III

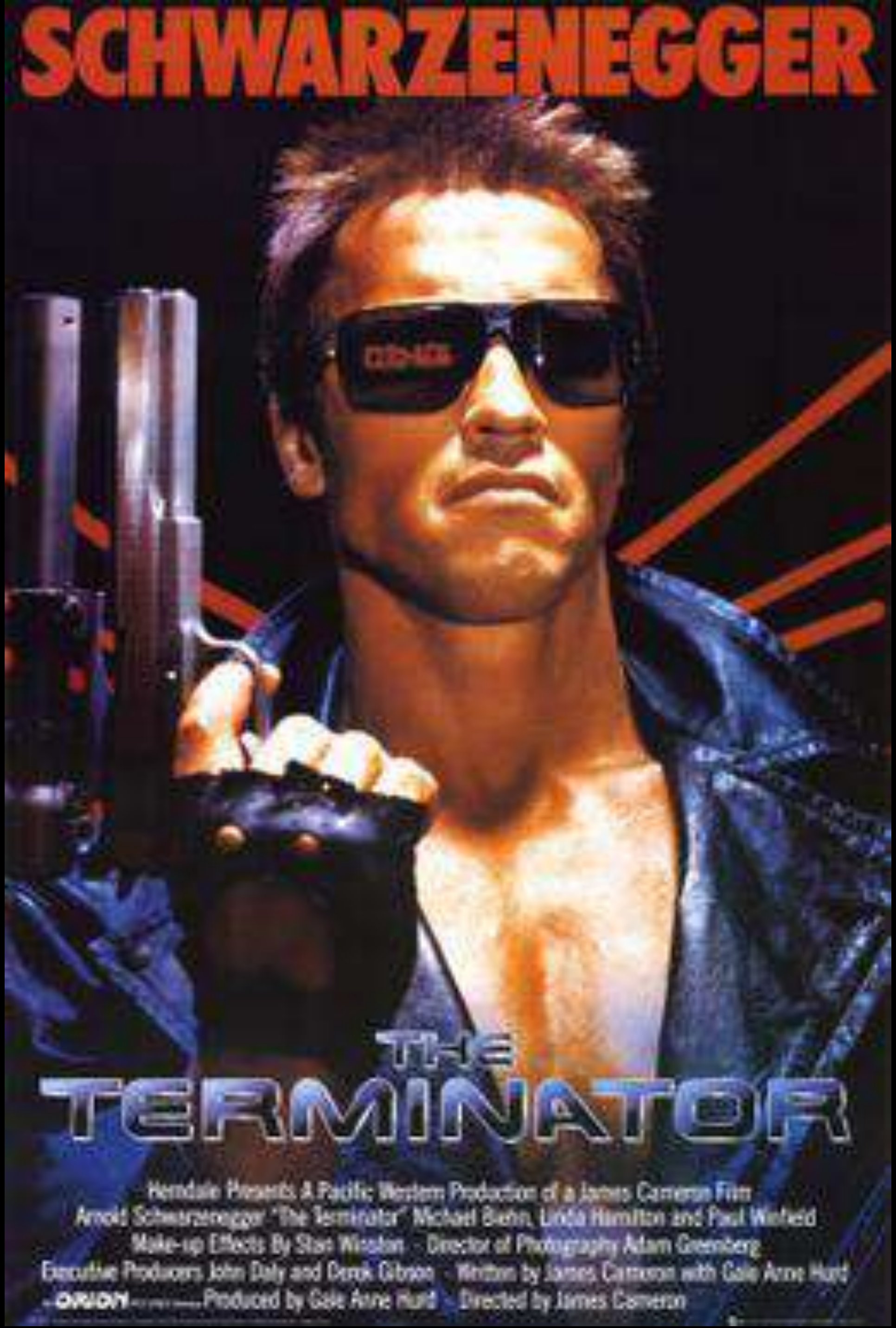
FORTUNE

TECH • ARTIFICIAL INTELLIGENCE

Let's Stop Freaking Out About Artificial Intelligence

BY ERIC SCHMIDT AND SEBASTIAN THRUN

June 28, 2016 7:23 PM GMT+2


SCHWARZENEGGER
THE
TERMINATOR
Hemdale Presents A Pacific Western Production of a James Cameron Film
Arnold Schwarzenegger "The Terminator" Michael Biehn, Linda Hamilton and Paul Winfield
Make-up Effects By Stan Winston - Director of Photography Adam Greenberg
Executive Producers John Daly and Derek Gibson - Written by James Cameron with Gale Anne Hurd
ORION Produced by Gale Anne Hurd - Directed by James Cameron

& INFORMATION VISIT
CARDIFFCITYFC.CO.UK
#CITYASONE
TICKET OFFICE
NO FACIAL
RECOGNITION
BIG BROTHER WATCH
DEFENDING CIVIL LIBERTIES, PROTECTING PRIVACY

#dragvsAI
AJL
PRESS PAUSE ON FACE SURVEILLANCE
Massachusetts
hat do you wa
Create?
MYHERO
Jean
Jean
Victor
Jean
Mison
ossie

ALGORITHM
WATCH

FIGHT FOR THE FUTURE

WHO ARE WE?

We are a group of artists, engineers, activists, and technologists who have been behind the largest online protests in human history, channeling Internet outrage into political power to win public interest victories previously thought to be impossible. We fight for a future where technology liberates — not oppresses — us. ✊

CAMPAIGN TO STOP
KILLER ROBOTS

www.sipeia.it

SIpEIA

Società Italiana per l'Etica dell'Intelligenza Artificiale