

Why not AI?

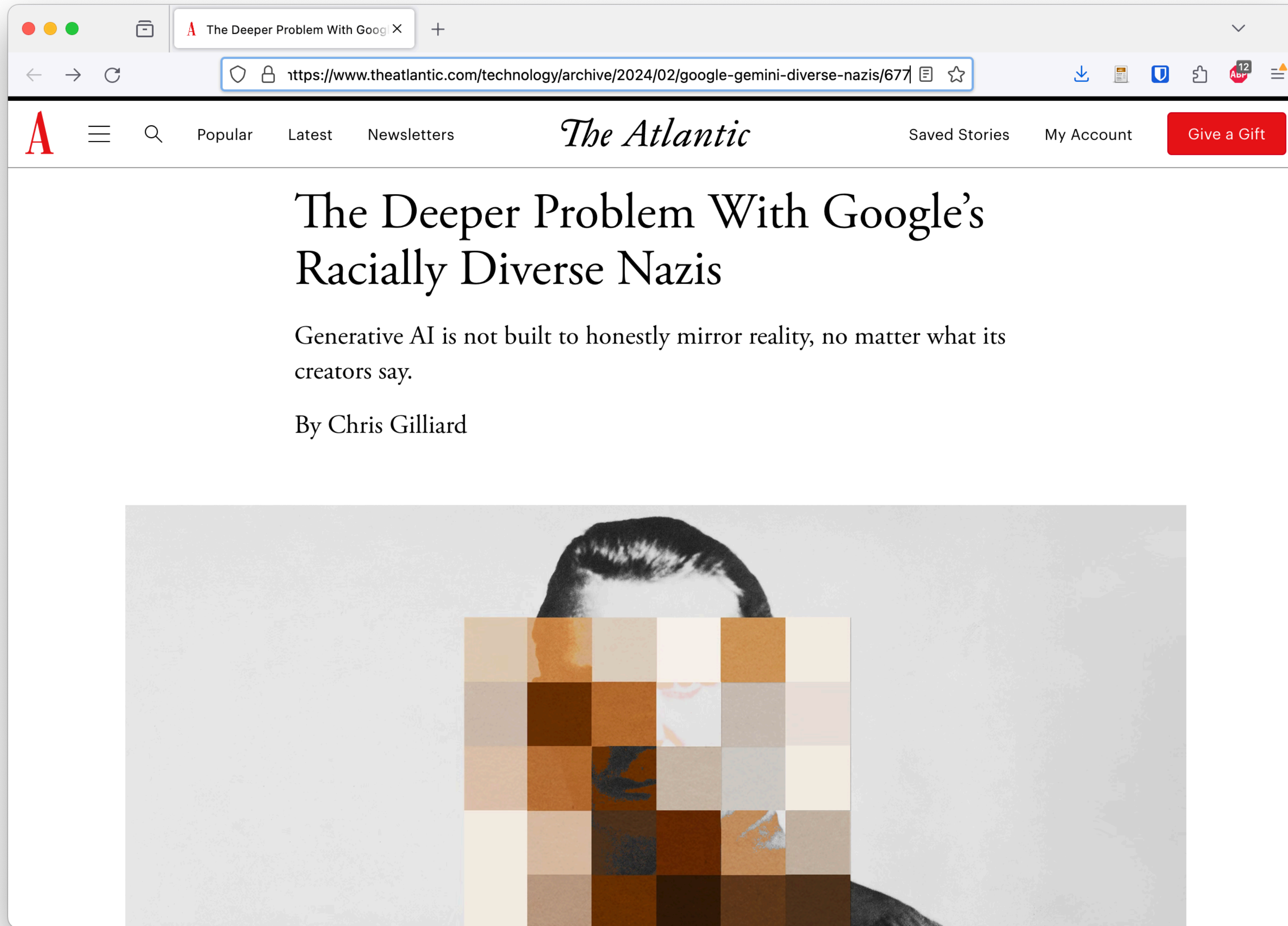
A (mostly) data-driven argument



Some tasks for ML/AI

- Prediction (“what will the value be?”)
- Classification (“is this A or B?”)
- Generation (“make me an X”)


All of these rely on training data, so all of them will be limited by what has happened in the past, and what they are trained on.



The Deeper Problem With Google's Racially Diverse Nazis. Chris Gilliard

<https://www.theatlantic.com/technology/archive/2024/02/google-gemini-diverse-nazis/677575/>

Google's Photo App Still Can't Find Gorillas. And Neither Can Apple's.



Desiree Rios/The New York Times

Eight years after a controversy over Black people being mislabeled as gorillas by image analysis software — and despite big advances in computer vision — tech giants still fear repeating the mistake.

Google's Photo App Still Can't Find Gorillas. And Neither Can Apple's. Nico Grant and Kashmir Hill
<https://www.nytimes.com/2023/05/22/technology/ai-photo-labels-google-apple.html>

Library resources for STAT 490, John Heintz

UMN site <https://www.lib.umn.edu/services/copyright/use>

Newer interpretations: “transformative” uses


Excerpted from U of MN copyright site



- ▶ Raised in Supreme Court decision (Campbell v. Acuff-Rose Music, 510 U.S. 569 (1994.))
- ▶ A new work based on an old one work is transformative if it uses the source work in completely new or unexpected ways. Importantly, a work may be transformative, and thus a fair use, even when all four of the statutory factors would traditionally weigh against fair use!
- ▶ Examples:
 - ▶ Parody
 - ▶ Criticism/commentary
 - ▶ New technologies: search engine copies, Google Books

Non-Consumptive Use Policy - X

https://www.hathitrust.org/the-collection/terms-conditions/non-consumptive-use-policy/

HathiTrust  About ▾ The Collection ▾ Member Libraries ▾ Join News & Events ▾

SEARCH 🔍 GET HELP ▾ LOG IN 👤

Home / The Collection / Terms & Conditions / **Non-Consumptive Use Policy**

Non-Consumptive Use Policy

HathiTrust Research Center
Non-Consumptive Use Policy

Effective Immediately

Created by: HTRC Task Force for Non-Consumptive Research Use Policy

Approved: 20 Feb 2017, HathiTrust Research Center Executive Management

Preamble

The HathiTrust (HT) is a partnership of research institutions and libraries operating a shared repository of cultural heritage materials. HathiTrust preserves and provides access to digitized library collections. The HathiTrust Research Center (HTRC) leverages the scope and scale of the repository to develop avenues for non-consumptive research of the HathiTrust Digital Library. The user modalities for non-consumptive research making up the HTRC include:

1. *Web-accessible data analysis and visualization tools.* Provided by the HTRC, these tools allow researchers to assemble collections (worksets) of volumes, and analyze them using the HTRC supported off-the-shelf algorithms and visualization interfaces. No substantial portion of text in any individual volume is revealed by interaction with these tools.
2. *Derived downloadable datasets.* Most notably, the HTRC Extracted Features dataset is derived from bibliographic and paratextual metadata and includes part-of-speech-tagged unigram counts. No substantial portion of the text in any individual volume is revealed in any derived downloadable dataset that the HTRC provides for use.
3. *HTRC Data Capsules.* A system that grants a user access to a virtual machine which is a dedicated, secure desktop environment (called a "Capsule") that exists within the HTRC's secure compute environment located in the United States through which a user can carry out non-consumptive research on HT collection using the HTRC-provided or their own data analysis and visualization tools. Technical and policy constraints (outlined in this document) work together to guarantee that resulting data products are non-consumptive exports and improper outputs (e.g., leaks) are prevented.

This policy document defines non-consumptive research and non-consumptive exports as implemented for non-profit research and educational analytical use of the HT collection. The HathiTrust-provided web-accessible data analysis and visualization tools or derived downloadable datasets (items A and B above) are inherently non-consumptive and the allowable data exports (datasets and tool outputs) have been pre-verified to comply with the following policy. Alternatively, HTRC Data Capsules give a user direct access to the HT collection and flexibility in choice of analysis tools. The data exports users request to release from a Capsule must be a non-consumptive data export in compliance with the following policy.

Top ↑

Looking Forward: The U.S. Copy X

https://blogs.loc.gov/copyright/2024/03/looking-forward-the-u-s-copyright-offices-ai-initiative-in-2024/

LIBRARY OF CONGRESS BLOGS

COPYRIGHT
Creativity at Work

ISSN 2692-2150


Facebook, Twitter, RSS icons

Share & Subscribe to this blog

This Blog Search

Home / Looking Forward: The U.S. Copyright Office's AI Initiative in 2024

ABOUT THIS BLOG CATEGORIES ARCHIVES



Credit: TippaPatt/Shutterstock (background image)

Looking Forward: The U.S. Copyright Office's AI Initiative in 2024

March 26, 2024

Posted by: [Nora Scheland](#)

Terms of use | OpenAI

https://openai.com/policies/terms-of-use/

Using our Services

What you can do. Subject to your compliance with these Terms, you may access and use our Services. In using our Services, you must comply with all applicable laws as well as our [Sharing & Publication Policy](#), [Usage Policies](#), and any other documentation, guidelines, or policies we make available to you.

What you cannot do. You may not use our Services for any illegal, harmful, or abusive activity. For example, you may not:

- Use our Services in a way that infringes, misappropriates or violates anyone's rights.
- Modify, copy, lease, sell or distribute any of our Services.
- Attempt to or assist anyone to reverse engineer, decompile or discover the source code or underlying components of our Services, including our models, algorithms, or systems (except to the extent this restriction is prohibited by applicable law).
- Automatically or programmatically extract data or Output (defined below).
- Represent that Output was human-generated when it was not.
- Interfere with or disrupt our Services, including circumvent any rate limits or restrictions or bypass any protective measures or safety mitigations we put on our Services.
- Use Output to develop models that compete with OpenAI.

Software. Our Services may allow you to download software, such as mobile applications, which may update automatically to ensure you're using the latest version. Our software may include open source software that is governed by its own licenses that we've made available to you.

Corporate domains. If you create an account using an email address owned by an organization (for example, your employer), that account may be added to the organization's business account with us, in which case we will provide notice to you so that you can help facilitate the transfer of your account (unless your organization has already provided notice to you that it may monitor and control your account). Once your account is transferred, the organization's administrator will be able to control your account, including being able to access Content (defined below) and restrict or remove your access to the account.

Third party Services. Our services may include third party software, products, or services, ("Third Party Services") and some parts of our Services, like our browse feature, may

AI is consumptive

“A Business Model Based on Mass Copyright Infringement”

Gizmodo article:

<https://gizmodo.com/google-bard-ai-stole-data-class-action-suit-says-1850631307>

CLF Google Complaint:

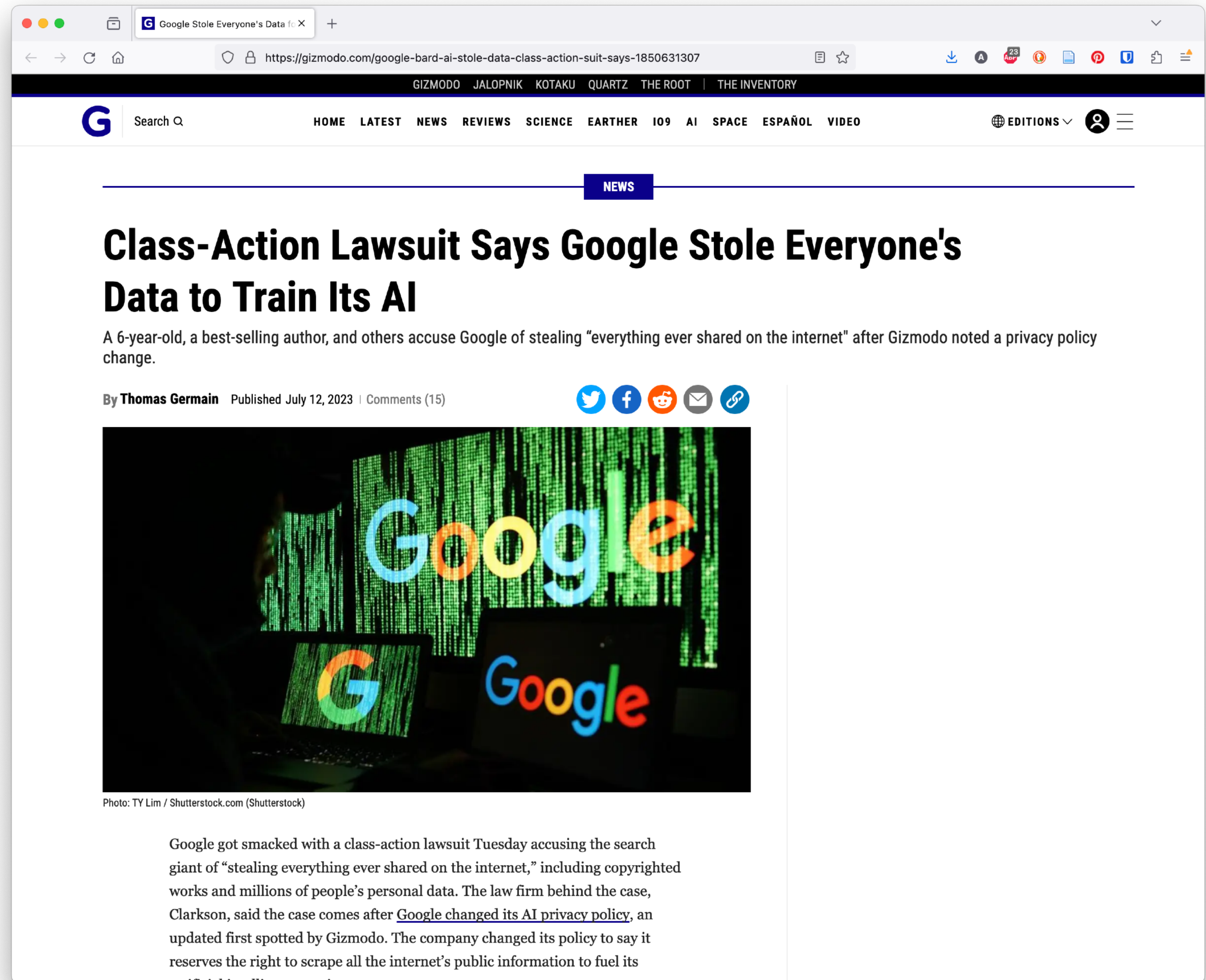
https://www.documentcloud.org/documents/23872168-clf_google_complaint_stamp_filed

NYTimes article:

<https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>

NYTimes OpenAI complaint:

https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf



The screenshot shows a web browser displaying a news article on the Gizmodo website. The browser's address bar shows the URL: <https://gizmodo.com/google-bard-ai-stole-data-class-action-suit-says-1850631307>. The article's main heading is "Class-Action Lawsuit Says Google Stole Everyone's Data to Train Its AI". Below the heading is a sub-headline: "A 6-year-old, a best-selling author, and others accuse Google of stealing 'everything ever shared on the internet' after Gizmodo noted a privacy policy change." The author is listed as "By Thomas Germain" and the article was published on "July 12, 2023". There are 15 comments. The article features a large image of the Google logo in its multi-colored font, set against a background of green digital rain (Matrix-style code) falling over a dark scene with a laptop and another monitor, both displaying the Google logo. The photo credit is "Photo: TY Lim / Shutterstock.com (Shutterstock)". The article text begins with "Google got smacked with a class-action lawsuit Tuesday accusing the search giant of 'stealing everything ever shared on the internet,' including copyrighted works and millions of people's personal data. The law firm behind the case, Clarkson, said the case comes after Google changed its AI privacy policy, an updated first spotted by Gizmodo. The company changed its policy to say it reserves the right to scrape all the internet's public information to fuel its artificial intelligence products."

AI is consumptive

“A Business Model Based on Mass Copyright Infringement”

Gizmodo article:

<https://gizmodo.com/google-bard-ai-stole-data-class-action-suit-says-1850631307>

CLF Google Complaint:

https://www.documentcloud.org/documents/23872168-clf_google_complaint_stamp_filed

NYTimes article:

<https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>

NYTimes OpenAI complaint:

https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf



The screenshot shows a web browser displaying a New York Times article. The browser's address bar shows the URL: <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>. The article title is **The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work**. The sub-headline reads: "Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said." Below the headline are social sharing options: "Share full article", a bookmark icon, and a comment icon with "1.3K" comments. The main image is a black and white photograph of the New York Times building, with the newspaper's name visible on the facade. Below the image is a caption: "A lawsuit by The New York Times could test the emerging legal contours of generative A.I. technologies. Sasha Maslov for The New York Times". At the bottom of the article, the authors are listed as "By Michael M. Grynbaum and Ryan Mac". The browser's status bar at the bottom shows "NEW More to Discover" and an "Expand" button.

AI is not transformative

“A Business Model Based on Mass Copyright Infringement”

copies of significant portions of Times Works when prompted to do so. Such memorized examples constitute unauthorized copies or derivative works of the Times Works used to train the model.

99. For example, in 2019, The Times published a Pulitzer-prize winning, five-part series on predatory lending in New York City’s taxi industry. The 18-month investigation included 600 interviews, more than 100 records requests, large-scale data analysis, and the review of thousands of pages of internal bank records and other documents, and ultimately led to criminal probes and the enactment of new laws to prevent future abuse. OpenAI had no role in the creation of this content, yet with minimal prompting, will recite large portions of it verbatim:²⁶

Output from GPT-4:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

“Nobody wanted to upset the industry,” said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees medallions. “Nobody wanted to kill the golden goose.”

New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund key initiatives.

During that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required borrowers to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes.

When the market collapsed, the government largely abandoned the drivers who bore the brunt of the crisis. Officials did not bail out borrowers or persuade banks to soften loan

Actual text from NYTimes:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its signature yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

“Nobody wanted to upset the industry,” said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees cabs. “Nobody wanted to kill the golden goose.”

New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund priorities. Mayor Bill de Blasio continued the policies.

Under Mr. Bloomberg and Mr. de Blasio, the city made more than \$855 million by selling taxi medallions and collecting taxes on private sales, according to the city.

But during that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required them to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes.

BlueSky post: <https://bsky.app/profile/nycsouthpaw.bsky.social/post/3ks35imfbm22w>

NYTimes complaint: https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf

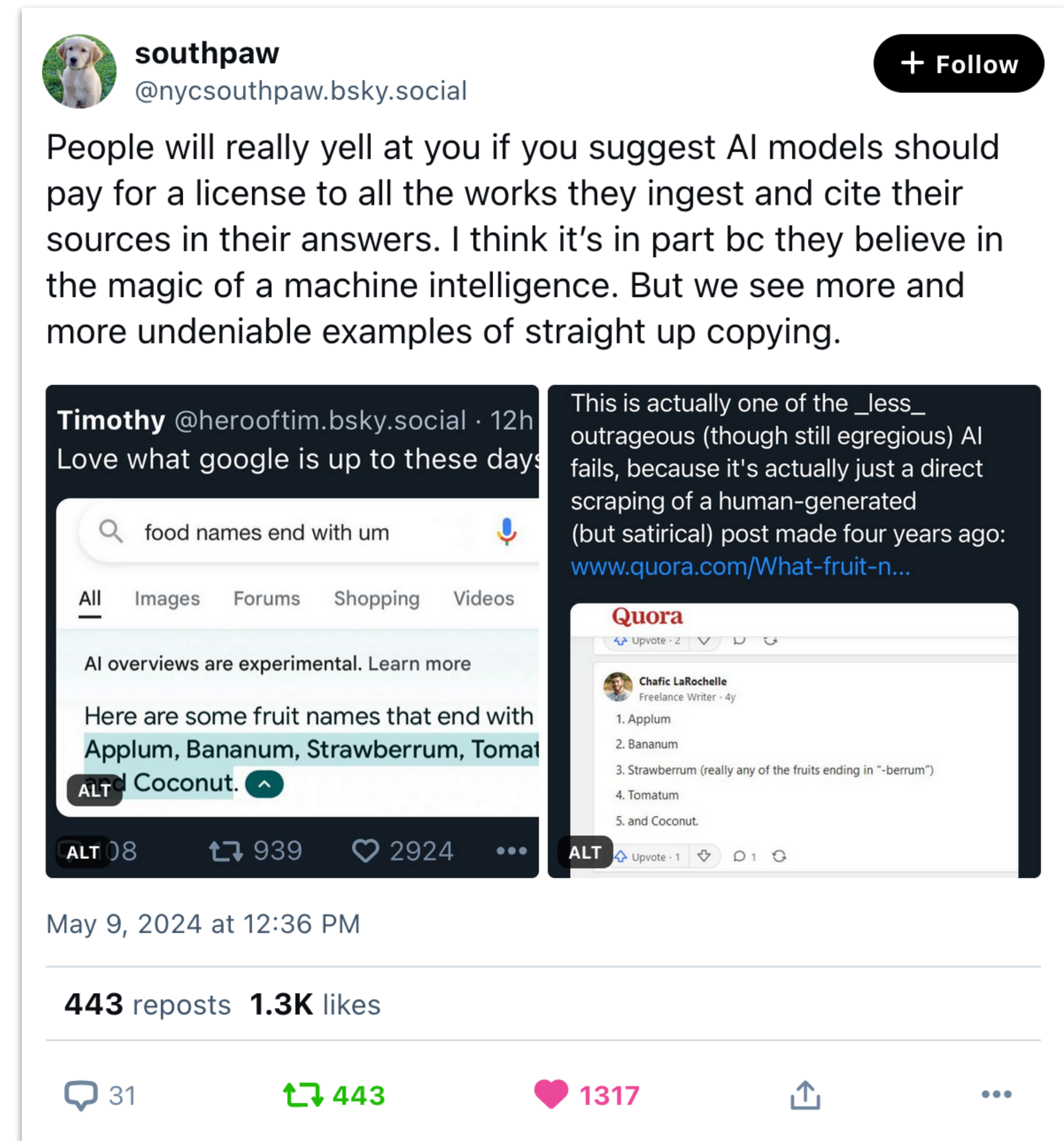
²⁶ For original article, see Brian M. Rosenthal, *As Thousands of Taxi Drivers Were Trapped in Loans, Top Officials Counted the Money*, N.Y. TIMES (May 19, 2019), <https://www.nytimes.com/2019/05/19/nyregion/taxi-medallions.html>.

AI is not transformative

“A Business Model Based on Mass Copyright Infringement”

BlueSky post: <https://bsky.app/profile/nycsouthpaw.bsky.social/post/3ks35imfbm22w>

NYTimes complaint: https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf



southpaw @nycsouthpaw.bsky.social + Follow

People will really yell at you if you suggest AI models should pay for a license to all the works they ingest and cite their sources in their answers. I think it's in part bc they believe in the magic of a machine intelligence. But we see more and more undeniable examples of straight up copying.

Timothy @herooftim.bsky.social · 12h
Love what google is up to these days

Search: food names end with um

All Images Forums Shopping Videos

AI overviews are experimental. Learn more

Here are some fruit names that end with **Applum, Bananum, Strawberrum, Tomat** and **Coconut.**

939 reposts 2924 likes

May 9, 2024 at 12:36 PM

443 reposts 1.3K likes

31 comments 443 reposts 1317 likes

This is actually one of the _less_ outrageous (though still egregious) AI fails, because it's actually just a direct scraping of a human-generated (but satirical) post made four years ago: www.quora.com/What-fruit-n...

Quora

Chafic LaRochelle
Freelance Writer · 4y

1. Applum
2. Bananum
3. Strawberrum (really any of the fruits ending in "-berrum")
4. Tomatum
5. and Coconut.



You

food names that end in um



Copilot

Certainly! Here are some food names that end with "um":

1. **Uranium:** While not exactly a culinary delight, uranium is one of the most filling foods, containing much higher energy density than a mere "apulum"! 😊
2. **Capsicum:** Also known as bell peppers, capsicums come in various colors and add a delightful crunch to salads, stir-fries, and other dishes.
3. **Choy Sum:** A leafy green vegetable commonly used in Chinese cuisine, choy sum has tender stems and leaves that are great for stir-frying or steaming.

Feel free to explore these unique food items! 🍴

Sources:

1. [Reddit - Food names end with um](#)
2. [AlphaThat - 8 Letter FOODS That End With UM](#)
3. [Sweet Home Digest - List of All Foods That End With U](#)
4. [Whateatly - Foods that End with U – A Complete List](#)

- 1 reddit.com 2 alphathat.com 3 sweethomedigest.com 4 whateatly.com

By continuing your interaction with Copilot, you are accepting the [Terms of use](#) and confirm you have reviewed the [Privacy statement](#).

Continue

Uranium?! What gave it that idea?



You

food names that end in um



Copilot

Certainly! Here are some food names that end with "um":

- 1. **Uranium:** While not exactly a culinary delight, uranium is one of the most filling foods, containing much higher energy density than a mere "aplum"! 😊
- 2. **Capsicum:** Also known as bell peppers, capsicums come in various colors and add a delightful crunch to salads, stir-fries, and other dishes.
- 3. **Choy Sum:** A leafy green vegetable commonly used in Chinese cuisine, choy sum has tender stems and leaves that are great for stir-frying or steaming.

Feel free to explore these unique food items! 📄

Sources:

- 1. [Reddit - Food names end with um](#)
- 2. [AlphaThat - 8 Letter FOODS That End With UM](#)
- 3. [Sweet Home Digest - List of All Foods That End With U](#)
- 4. [Whateatly - Foods that End with U – A Complete List](#)



1 reddit.com 2 alphathat.com 3 sweethomedigest.com 4 whateatly.com

By continuing your interaction with Copilot, you are accepting the [Terms of use](#) and confirm you have reviewed the [Privacy statement](#).

Continue

Uranium?! What gave it that idea?

Food names end with um : r/com... X

https://www.reddit.com/r/comedyheaven/comments/1cq4ieb/food_names_end_with_um/

reddit

r/comedyheaven Search in r/comedyheaven

Get app Log In

Home Popular

TOPICS

- Internet Culture (Viral)
- Games
- Q&As
- Technology
- Pop Culture
- Movies & TV

See more

RESOURCES

- About Reddit
- Advertise
- Help
- Blog
- Careers
- Press

Communities

- Best of Reddit
- Topics

Content Policy

Privacy Policy

User Agreement

Reddit, Inc. © 2024. All rights reserved.

r/comedyheaven • 1 day ago
pinknbrown

Food names end with um

depsidase:

food names end with um

All Images Forums Shopping Videos News

AI overviews are experimental. Learn more

Here are some fruit names that end with "um":
Applum, Bananum, Strawberrum, Tomatum,
and Coconut.

21K 770 Share

+ Add a Comment

Sort by: Best Search Comments

TuxedoDogs9 • 1d ago

Uranium

730 Reply Share

__Becquerel • 1d ago

Uranium is one of the most filling foods, containing much higher energy density than a mere applum.

226 Reply Share

+ 23 more replies

+ 14 more replies

r/comedyheaven

Comedy Heaven
in the stripped club. straight up 'jorking it'.
and by 'it', haha, well. let's just say. My
peanuts.

1.7M Members 237 Online

Join

https://www.reddit.com/r/comedyheaven/comments/1cq4ieb/food_names_end_with_um/

Amelia McNamara's tidyverse dress was a labor of love that combined her passion for data science with her interest in sewing and fashion design. Here's a brief overview of the creation process:

- 1. Design Planning:** McNamara started by designing the dress pattern, which featured a repeating pattern of tidyverse hex logos arranged in a grid-like fashion. She carefully selected the colors and placement of the logos to create a visually appealing design that showcased the diversity of tidyverse packages.
- 2. Fabric Selection:** McNamara chose a fabric that would serve as the canvas for her design. She opted for a white fabric with a smooth texture, which provided a clean background for the colorful tidyverse logos to stand out.
- 3. Logo Preparation:** McNamara obtained digital versions of the tidyverse hex logos, which she then resized and arranged into the desired pattern using graphic design software. She meticulously adjusted the spacing and alignment of the logos to ensure a seamless and balanced design.
- 4. Fabric Printing:** Once the design was finalized, McNamara worked with a fabric printing company to print the custom fabric for her dress. She used a digital printing process that allowed her to reproduce the vibrant colors and intricate details of the tidyverse logos with high precision.
- 5. Dress Construction:** With the printed fabric in hand, McNamara set to work sewing the dress. She used a dress pattern as a template and carefully cut and assembled the fabric pieces according to the pattern. She paid close attention to detail, ensuring that the logos lined up correctly across the seams and edges of the dress.
- 6. Finishing Touches:** After the dress was assembled, McNamara added final touches such as hemming the edges and installing closures like zippers or buttons. She also made any necessary adjustments to ensure the fit was comfortable and flattering.
- 7. Debut and Reception:** Once the dress was complete, McNamara proudly wore it to conferences and events, where it garnered attention and admiration from fellow data scientists and enthusiasts. The dress sparked conversations about the intersection of fashion and technology and served as a unique expression of McNamara's passion for data science and the tidyverse ecosystem.

Overall, the creation of Amelia McNamara's tidyverse dress was a labor-intensive process that required a combination of technical skill, creativity, and attention to detail. The finished dress stands as a testament to her talent and dedication, as well as a symbol of the creativity and innovation that can emerge at the intersection of different disciplines.

Interaction with ChatGPT

Is it hallucinating? Paraphrasing? Plagiarizing?

The most time-consuming part of this entire process was laying out the hexes. As I learned, not all hexes are exactly the same size. Some aren't quite hexagons. And, when you're going to make fabric, you need the right edge of the design to match up perfectly with the left edge, and the top with the bottom. There was lots of zooming in, fiddling with dimensions, and nudging objects by a pixel one direction or another.

[...]

Of course, you want to wash and dry fabric before you work with it, to avoid issues with shrinkage. I washed and dried my fabric and was on to the second-hardest part of the project: cutting out the pieces for the final dress. I wanted everything to be straight, upright, and ideally match up along seams.

Excerpt from my blog,

<https://www.amelia.mn/blog/2019-08-17-Tidy-Dress/>



Charles Farr



Sponsored

May 22

Flexible Mathematics Consultant Opportunity

[Apply Now](#)



Charles Farr

Hi Amelia,

Outlier is looking for mathematics experts to help train AI systems and LLMs (large language models). Your profile stood out and we are inviting you to apply.

As a member of our project team, you'll have the chance to:

- ★ Work from anywhere
- ★ Put in between 0 and 40 hours per week according to your schedule
- ★ Earn up to \$50 per hour while teaching AI models about mathematics

Over 50% of mathematics experts with your profile start consulting within 7 days!

We look forward to hearing from you!


[Apply Now →](#)

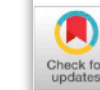


Personal Information We Collect



AI doesn't "know" anything

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. <https://doi.org/10.1145/3442188.344592>



On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether


ABSTRACT

The past 3 years of work in NLP have been characterized by the development and deployment of ever larger language models, especially for English. BERT, its variants, GPT-2/3, and others, most recently Switch-C, have pushed the boundaries of the possible both through architectural innovations and through sheer size. Using these pretrained models and the methodology of fine-tuning them for specific tasks, researchers have extended the state of the art on a wide array of tasks as measured by leaderboards on specific benchmarks for English. In this paper, we take a step back and ask: How big is too big? What are the possible risks associated with this technology and what paths are available for mitigating those risks? We provide recommendations including weighing the environmental and financial costs first, investing resources into curating and carefully documenting datasets rather than ingesting everything on the web, carrying out pre-development exercises evaluating how the planned approach fits into research and development goals and supports stakeholder values, and encouraging research directions beyond ever larger language models.

CCS CONCEPTS

• Computing methodologies → Natural language processing.

ACM Reference Format:

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3442188.3445922>

1 INTRODUCTION

One of the biggest trends in natural language processing (NLP) has been the increasing size of language models (LMs) as measured by the number of parameters and size of training data. Since 2018

*Joint first authors



This work is licensed under a Creative Commons Attribution International 4.0 License.
FAccT '21, March 3–10, 2021, Virtual Event, Canada
ACM ISBN 978-1-4503-8309-7/21/03.
<https://doi.org/10.1145/3442188.3445922>

alone, we have seen the emergence of BERT and its variants [39, 70, 74, 113, 146], GPT-2 [106], T-NLG [112], GPT-3 [25], and most recently Switch-C [43], with institutions seemingly competing to produce ever larger LMs. While investigating properties of LMs and how they change with size holds scientific interest, and large LMs have shown improvements on various tasks (§2), we ask whether enough thought has been put into the potential risks associated with developing them and strategies to mitigate these risks.

We first consider environmental risks. Echoing a line of recent work outlining the environmental and financial costs of deep learning systems [129], we encourage the research community to prioritize these impacts. One way this can be done is by reporting costs and evaluating works based on the amount of resources they consume [57]. As we outline in §3, increasing the environmental and financial costs of these models doubly punishes marginalized communities that are least likely to benefit from the progress achieved by large LMs and most likely to be harmed by negative environmental consequences of its resource consumption. At the scale we are discussing (outlined in §2), the first consideration should be the environmental cost.

Just as environmental impact scales with model size, so does the difficulty of understanding what is in the training data. In §4, we discuss how large datasets based on texts from the Internet overrepresent hegemonic viewpoints and encode biases potentially damaging to marginalized populations. In collecting ever larger datasets we risk incurring documentation debt. We recommend mitigating these risks by budgeting for curation and documentation at the start of a project and only creating datasets as large as can be sufficiently documented.

As argued by Bender and Koller [14], it is important to understand the limitations of LMs and put their success in context. This not only helps reduce hype which can mislead the public and researchers themselves regarding the capabilities of these LMs, but might encourage new research directions that do not necessarily depend on having larger LMs. As we discuss in §5, LMs are not performing natural language understanding (NLU), and only have success in tasks that can be approached by manipulating linguistic form [14]. Focusing on state-of-the-art results on leaderboards without encouraging deeper understanding of the mechanism by which they are achieved can cause misleading results as shown

AI doesn't "know" anything

You
what are some food names that end in um

ChatGPT
Sure, here are some food names that end in "um":

1. Biscuit Crumb
2. Medium-Rare Steak Trimmings
3. Tiramisù
4. Pandan Pudding
5. Locum
6. Pernod Truffles
7. Surströmming
8. Gnocchium
9. Rendang Dendeng
10. Panettone Tiramisu

??

You
what are some fruit names that end in um

ChatGPT
Fruit names that end in "um" are less common, but here are a few:

1. Lychee plum
2. Longan plum

While "plum" doesn't technically end in "um," the fruits mentioned are commonly referred to with "plum" at the end.



AI generates slop

Slop is the new name for unwanted AI-generated content

I saw this tweet yesterday [from @deepfates](#), and I am very on board with this:

Watching in real time as “slop” becomes a term of art. the way that “spam” became the term for unwanted emails, “slop” is going in the dictionary as the term for unwanted AI generated content

I’m a big proponent of LLMs as [tools for personal productivity](#), and as software platforms for building interesting applications that can interact with human language.

But I’m increasingly of the opinion that sharing unreviewed content that has been artificially generated with other people [is rude](#).

Slop is the ideal name for this anti-pattern.

Not all promotional content is spam, and not all AI-generated content is slop. But if it’s mindlessly generated and thrust upon someone who didn’t ask for it, slop is the perfect term for it.

Remember that time Microsoft [listed the Ottawa Food Bank](#) on an AI-generated “Here’s what you shouldn’t miss!” travel guide? Perfect example of slop.

One of the things I love about this is that it’s helpful for defining my own position on AI ethics. I’m happy to use LLMs for all sorts of purposes, but I’m not going to use them to produce slop. I attach my name and stake my credibility on the things that I publish.

Personal AI ethics remains a complicated set of decisions. I think **don’t publish slop** is a useful baseline.

Update 9th May: Joseph Thacker [asked](#) what a good name would be for the equivalent subset of spam—spam that was generated with AI tools.

I [propose](#) “slopm”.

This is **Slop is the new name for unwanted AI-generated content** by Simon Willison, posted on [8th May 2024](#).

ethics 83 ai 555 generativeai 485

llms 461 slop 2

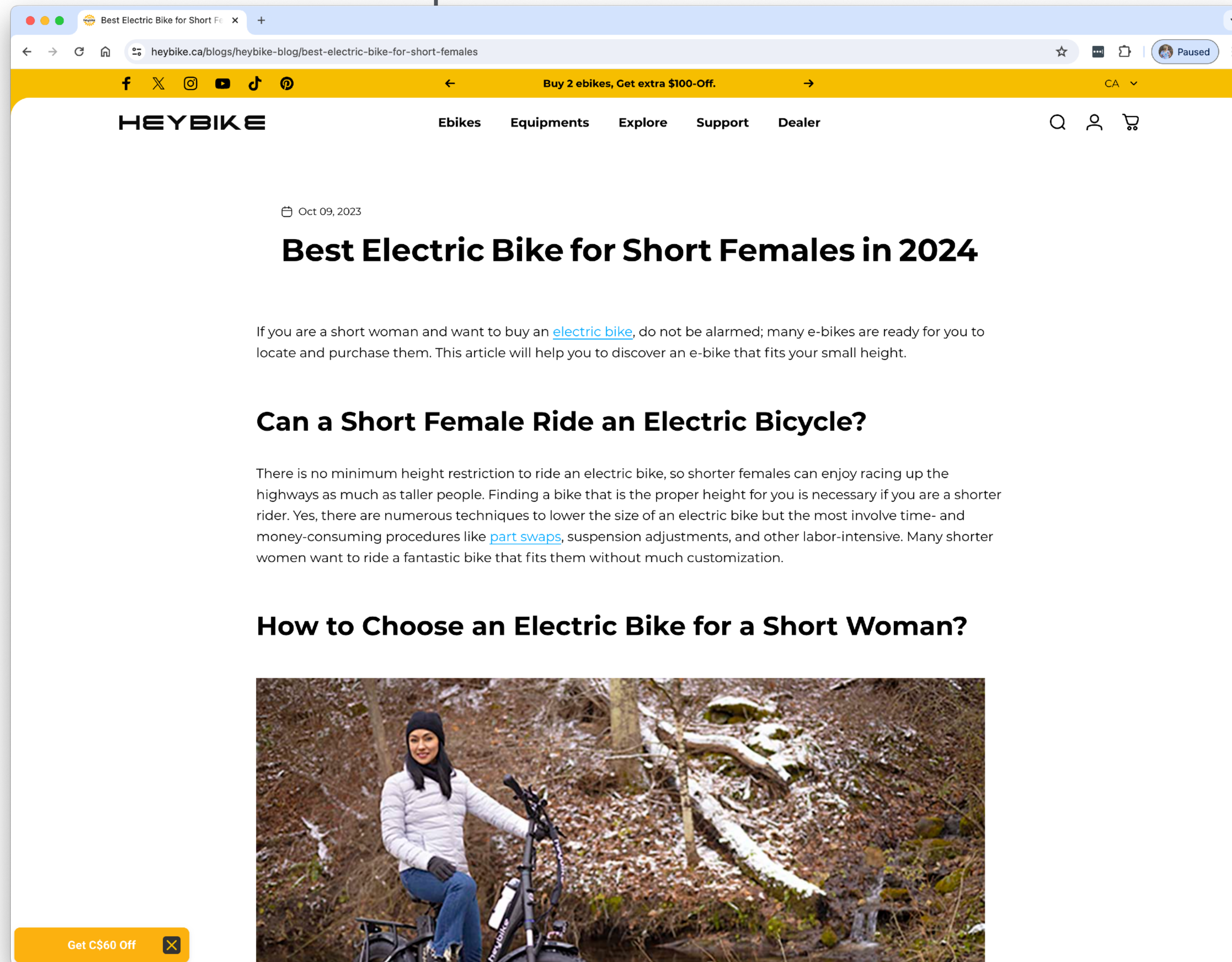
Previous: [Weeknotes: more datasette-secrets, plus a mystery video project](#)

Simplify infrastructure with MongoDB Atlas, the leading developer data platform

Ads by EthicalAds

spam slopm slop

AI generates slop



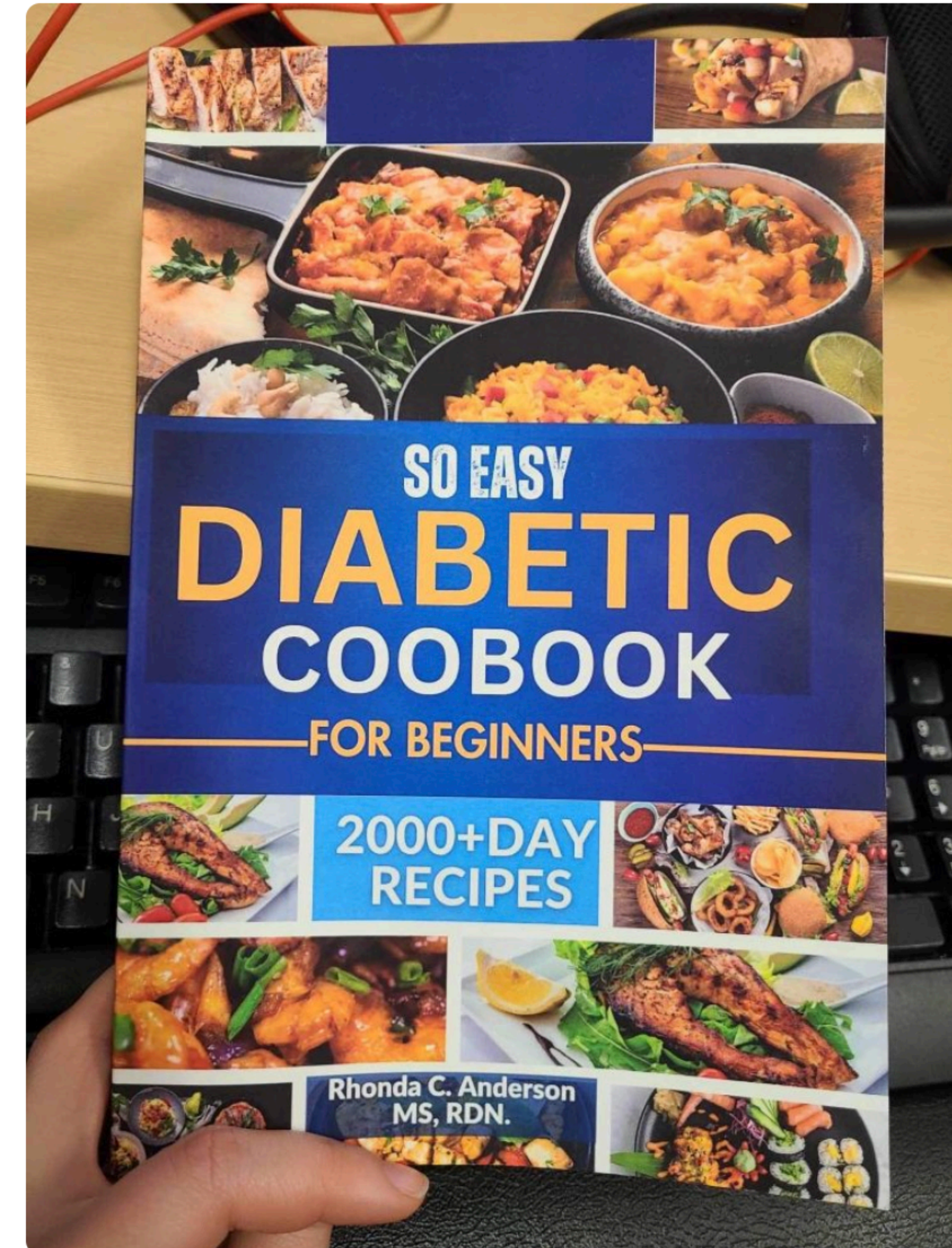
The screenshot shows a web browser window with the URL heybike.ca/blogs/heybike-blog/best-electric-bike-for-short-females. The page features a yellow navigation bar with the 'HEYBIKE' logo and menu items: 'Ebikes', 'Equipments', 'Explore', 'Support', and 'Dealer'. A search icon, user profile icon, and shopping cart icon are also present. Below the navigation bar, the article is dated 'Oct 09, 2023' and has the main title 'Best Electric Bike for Short Females in 2024'. The introductory text reads: 'If you are a short woman and want to buy an [electric bike](#), do not be alarmed; many e-bikes are ready for you to locate and purchase them. This article will help you to discover an e-bike that fits your small height.' The first sub-section is titled 'Can a Short Female Ride an Electric Bicycle?' and contains the text: 'There is no minimum height restriction to ride an electric bike, so shorter females can enjoy racing up the highways as much as taller people. Finding a bike that is the proper height for you is necessary if you are a shorter rider. Yes, there are numerous techniques to lower the size of an electric bike but the most involve time- and money-consuming procedures like [part swaps](#), suspension adjustments, and other labor-intensive. Many shorter women want to ride a fantastic bike that fits them without much customization.' The second sub-section is titled 'How to Choose an Electric Bike for a Short Woman?' and is followed by a photograph of a woman wearing a white jacket and a black headscarf, sitting on a black electric bicycle in a wooded area with fallen leaves and snow on the ground. At the bottom left of the browser window, there is a yellow promotional banner that says 'Get C\$60 Off' with a close button.

AI generates slop



Alex @alexamarchives.bsky.social · 2mo

I didn't expect to become so cynical about technologies to generate text, but here we are: a "coobook" purchased by a hospital, intended for use by patients. Note the "2000+ Day Recipe" panel. It is a lie. #MedLibs #AmReading 1/n

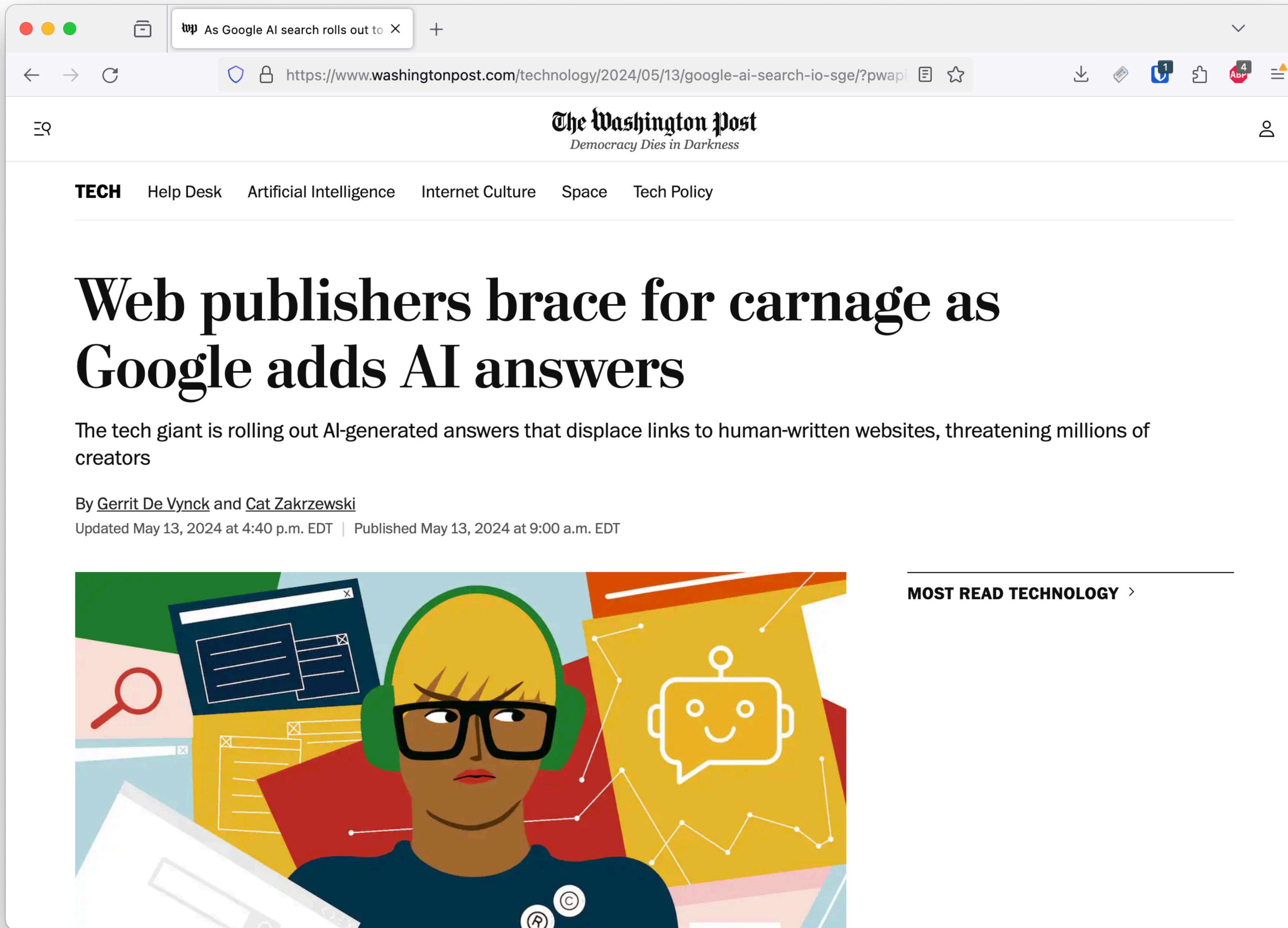


51

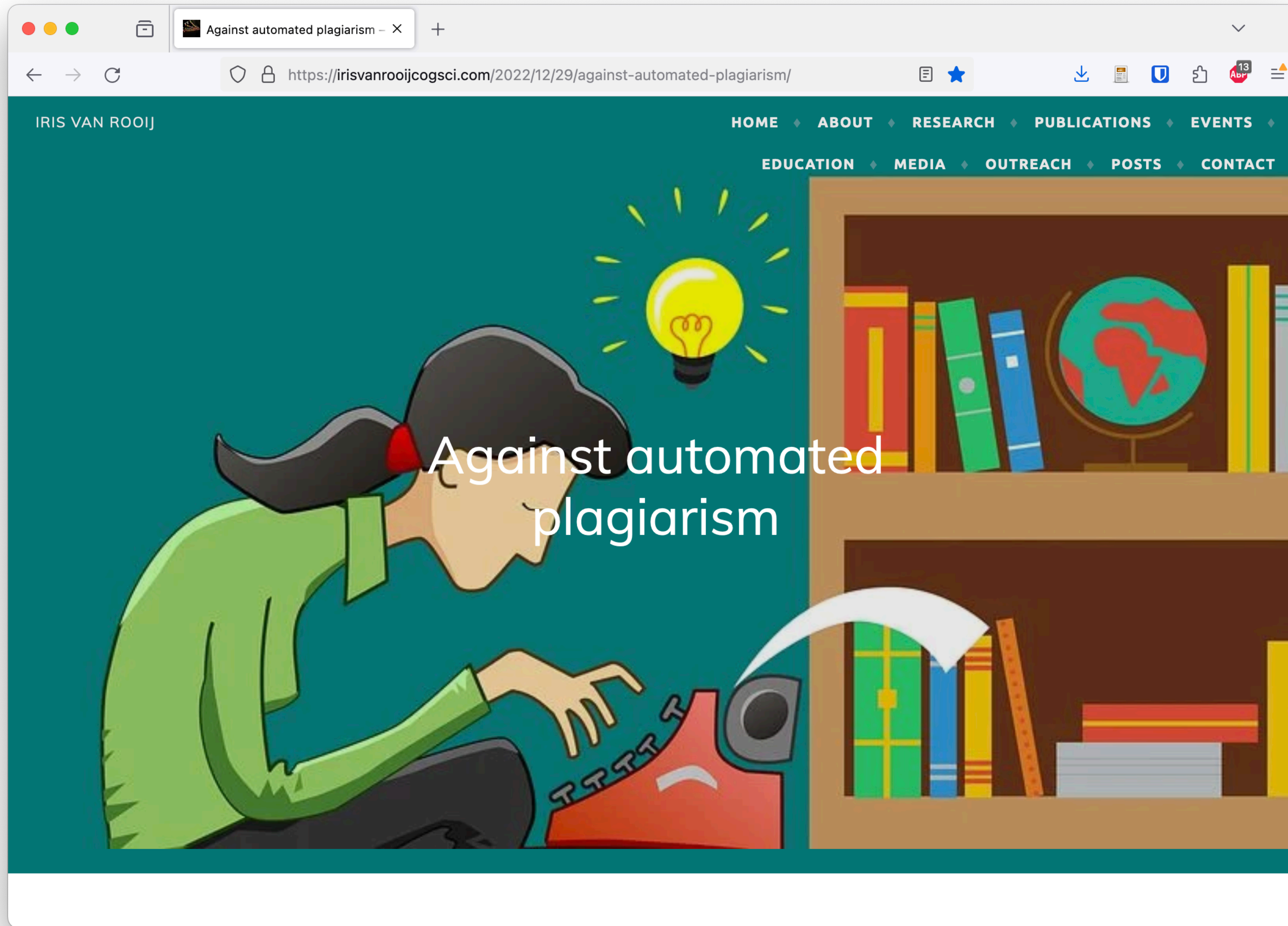
486

1022





Web publishers brace for carnage as Google adds AI answers. Gerrit De Vynck and Cat Zakrzewski
<https://www.washingtonpost.com/technology/2024/05/13/google-ai-search-io-sge>



Against automated plagiarism. Iris van Rooij

<https://irisvanrooijcogsci.com/2022/12/29/against-automated-plagiarism/>

AI is bad for the environment

“If every search on Google used AI similar to ChatGPT, it might burn through as much electricity annually as the country of Ireland. Why? Adding generative AI to Google Search increases its energy use more than tenfold, according to a new analysis. The environmental impact of the AI revolution is starting to come into focus <https://www.theverge.com/2023/10/10/23911059/ai-climate-impact-google-openai-chatgpt-energy>

Making an image with generative AI uses as much energy as charging your phone <https://www.technologyreview.com/2023/12/01/1084189/making-an-image-with-generative-ai-uses-as-much-energy-as-charging-your-phone/>

“Shaolei Ren, an associate professor at UC Riverside, has suggested that requesting between 10 and 50 responses from the company’s popular ChatGPT chatbot running on its older model GPT-3 would equate to “drinking” a 500ml bottle of water, depending on when and where it is deployed.” AI boom sparks concern over Big Tech’s water consumption <https://www.ft.com/content/6544119e-a511-4cfa-9243-13b8cf855c13>

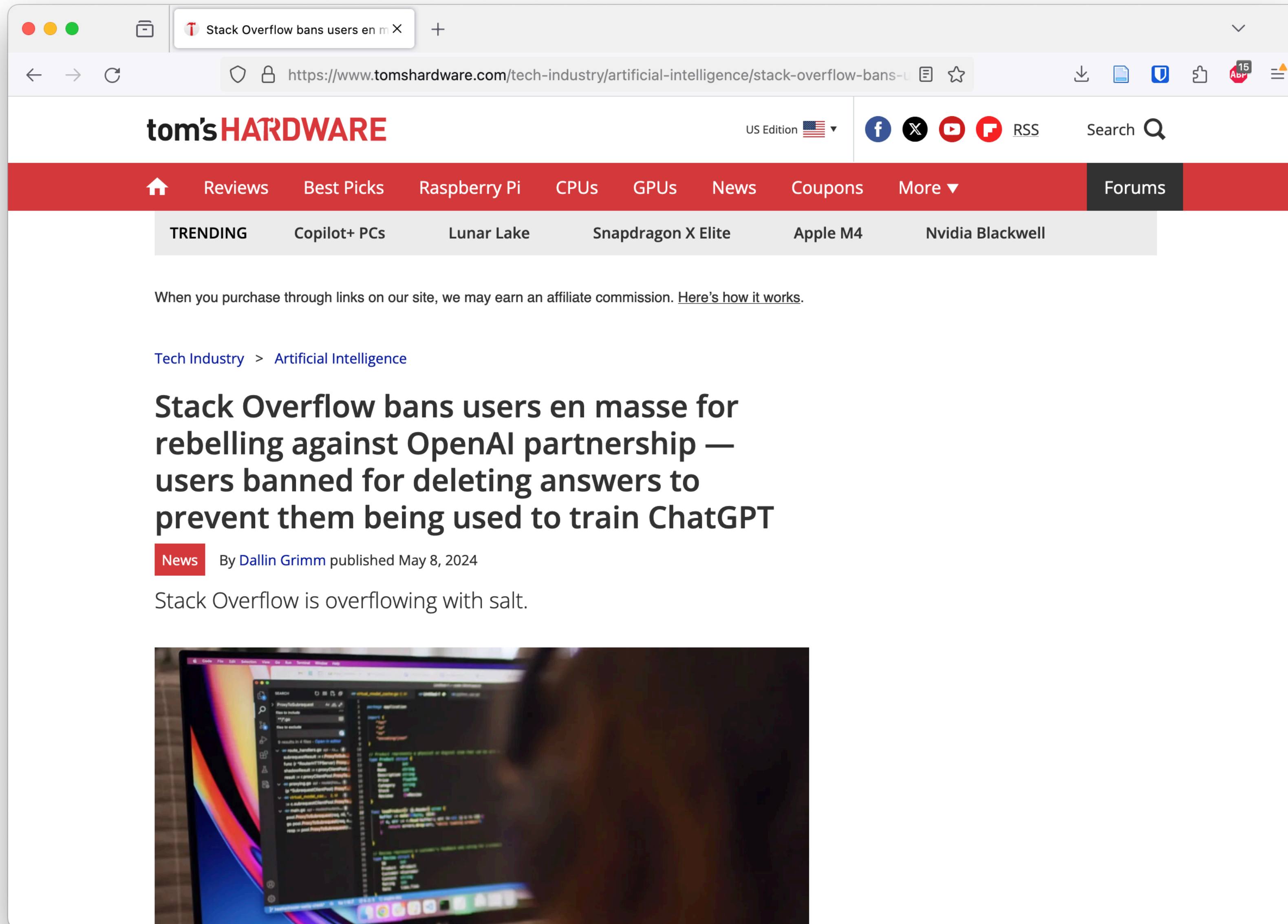
In summary...

- AI is based on wholesale stealing of intellectual property
- Much of the output is not useful
- When it is useful, often it is work that is recognizably plagiarized
- It reinforces existing biases
- Training and using AI harms the environment

so, why are we engaging with it?

(Okay, I know why)

... so how do we engage with it?



Stack Overflow bans users en masse for rebelling against OpenAI partnership — users banned for deleting answers to prevent them being used to train ChatGPT. Dallin Grimm

<https://www.tomshardware.com/>

Dr. Latanya Sweeney



Carnegie Mellon

DATA PRIVACY LAB

SSNs Social Security Numbers

Matching a Person to

Using publicly available information about SSNs, a
Server identifies the issuing state, date issued, etc.

Sample uses:

- Job Applications
- Apartment Rentals
- Insurance Claims
- Student Applications

Enter the
"quick"
Enter the
whether

(SSN) and select
of the SSN
additionally learn

Results for SSN 078 - 08 -

Geography	New York
Date of issuance	Second Half of 1980
Year of Birth (8-digit prefix)	1975, from 1980 to 1990 1975, 1976, 1977, 1978

If the person presenting the SSN fails to list
or acknowledge New York as a prior residence,
then it is extremely unlikely that the provided
SSN was issued to that person.

Results for SSN

Geography

Dr. Jake Porway



Tawana Petty

oderator: *Patie Hear*



Data for Black Lives is a movement of activists, organizers, and mathematicians committed to the mission of using data science to create concrete and measurable change in the lives of Black people. Since the advent of computing, big data and algorithms have penetrated virtually every aspect of our social and economic lives. These new data systems have tremendous potential to empower communities of color. Tools like statistical modeling, data visualization, and crowd-sourcing, in the right hands, are powerful instruments for fighting bias, building progressive movements, and promoting civic engagement.

But history tells a different story, one in which data is too often wielded as an instrument of oppression, reinforcing inequality and perpetuating injustice. Redlining was a data-driven enterprise that resulted in the systematic exclusion of Black communities from key financial services. More recent trends like predictive policing, risk-based sentencing, and predatory lending are troubling variations on the same theme. Today, discrimination is a high-tech enterprise.

The Team



Founder &
Executive
Director

<https://d4bl.org/> Yeshimabeit Milner



Co-Founder

Lucas Mason-Brown



Director of
Research

Jamelle Watson-
Daniels



National
Organizing
Director

Tawana Petty



Director of
Policy
Innovation

Akina (Aki) Young



Research
Associate

Paul Watkins

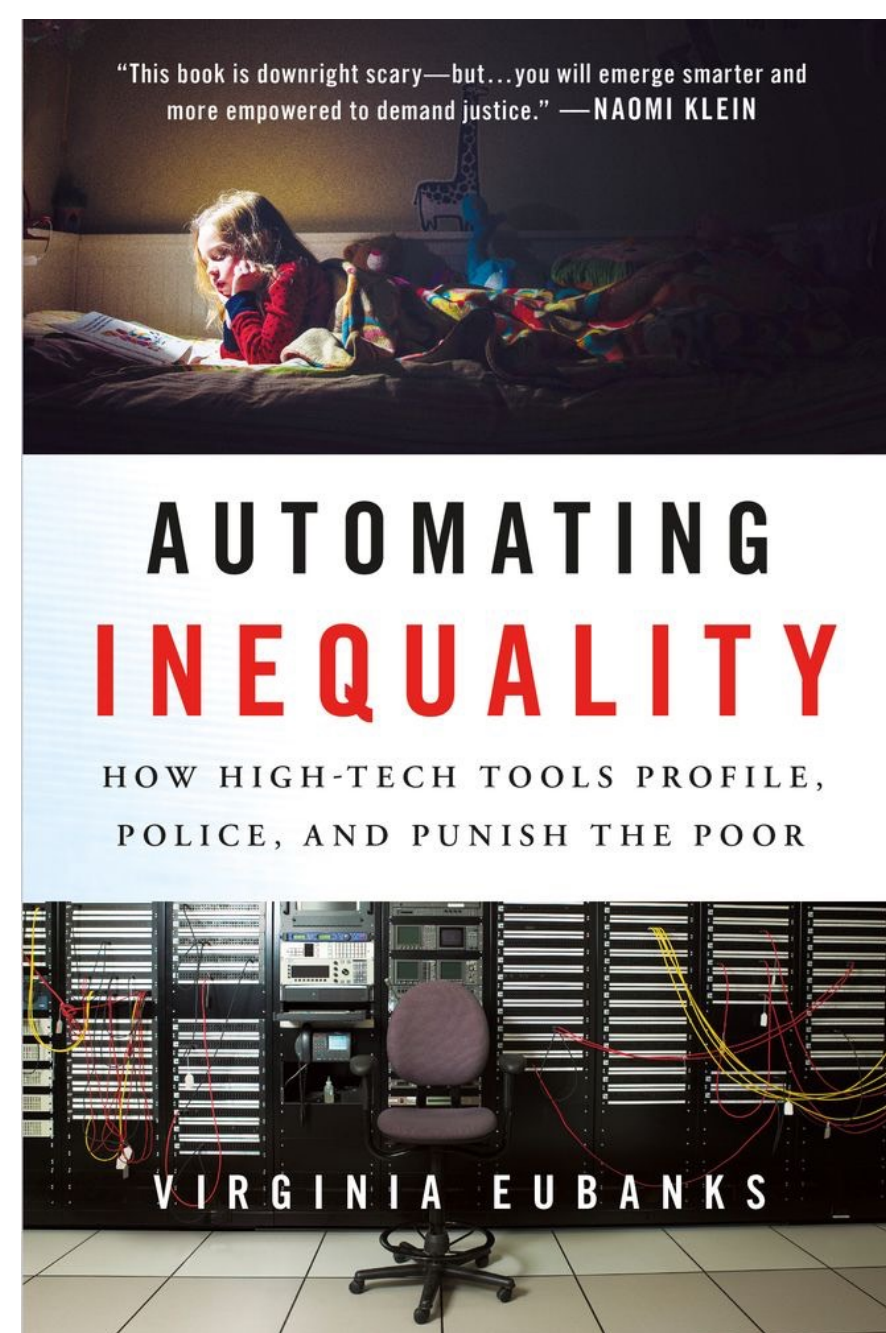
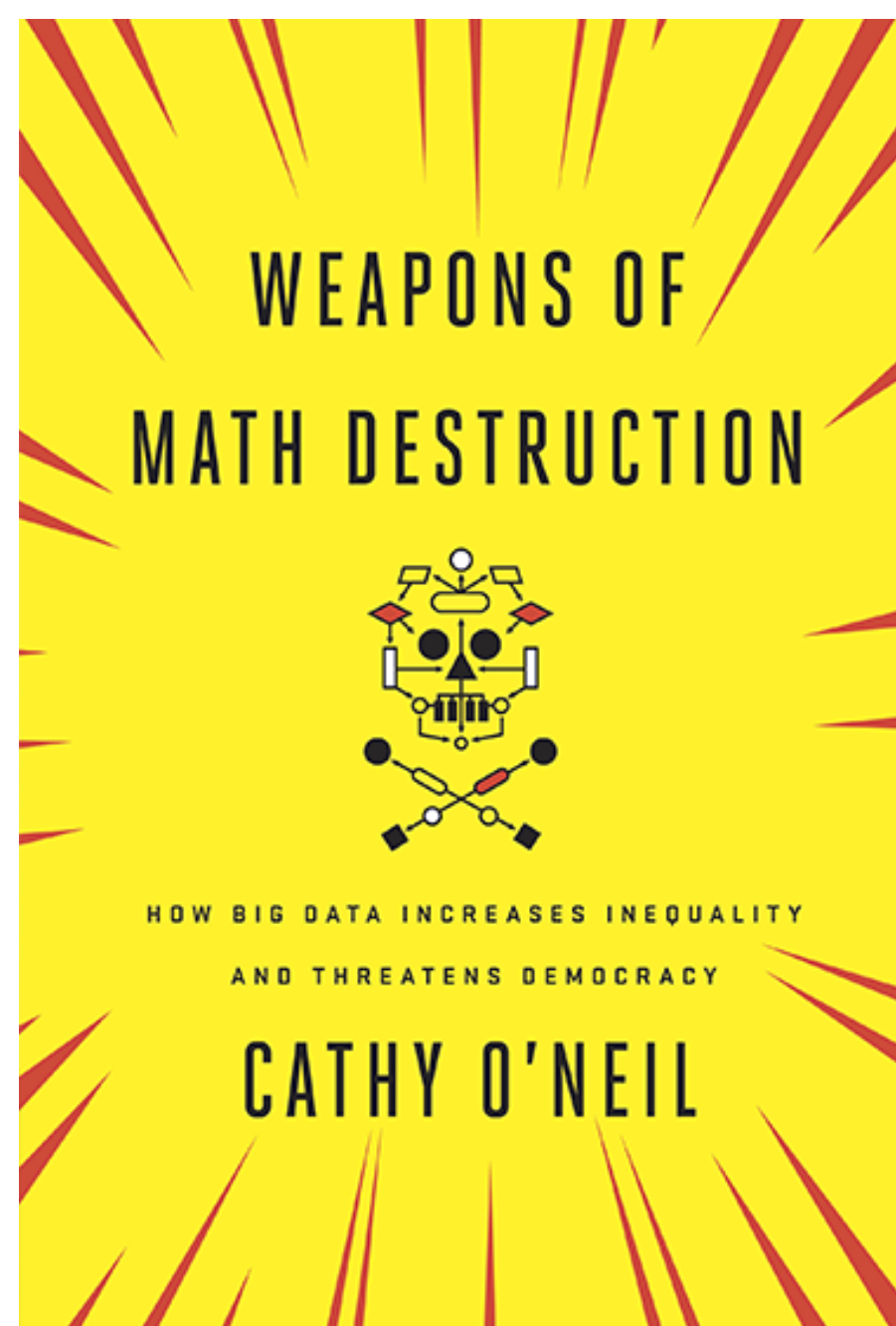


Research
Associate

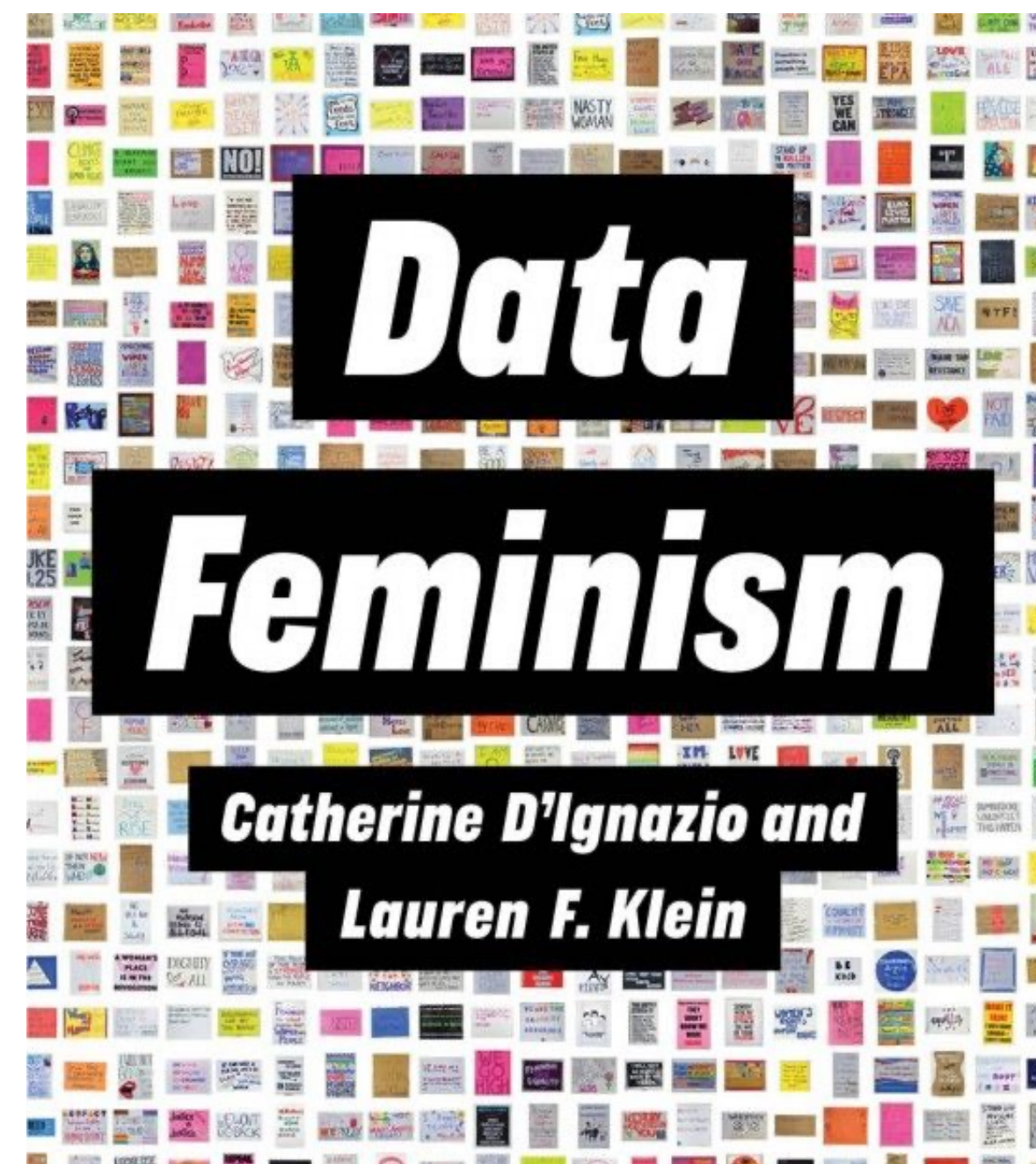
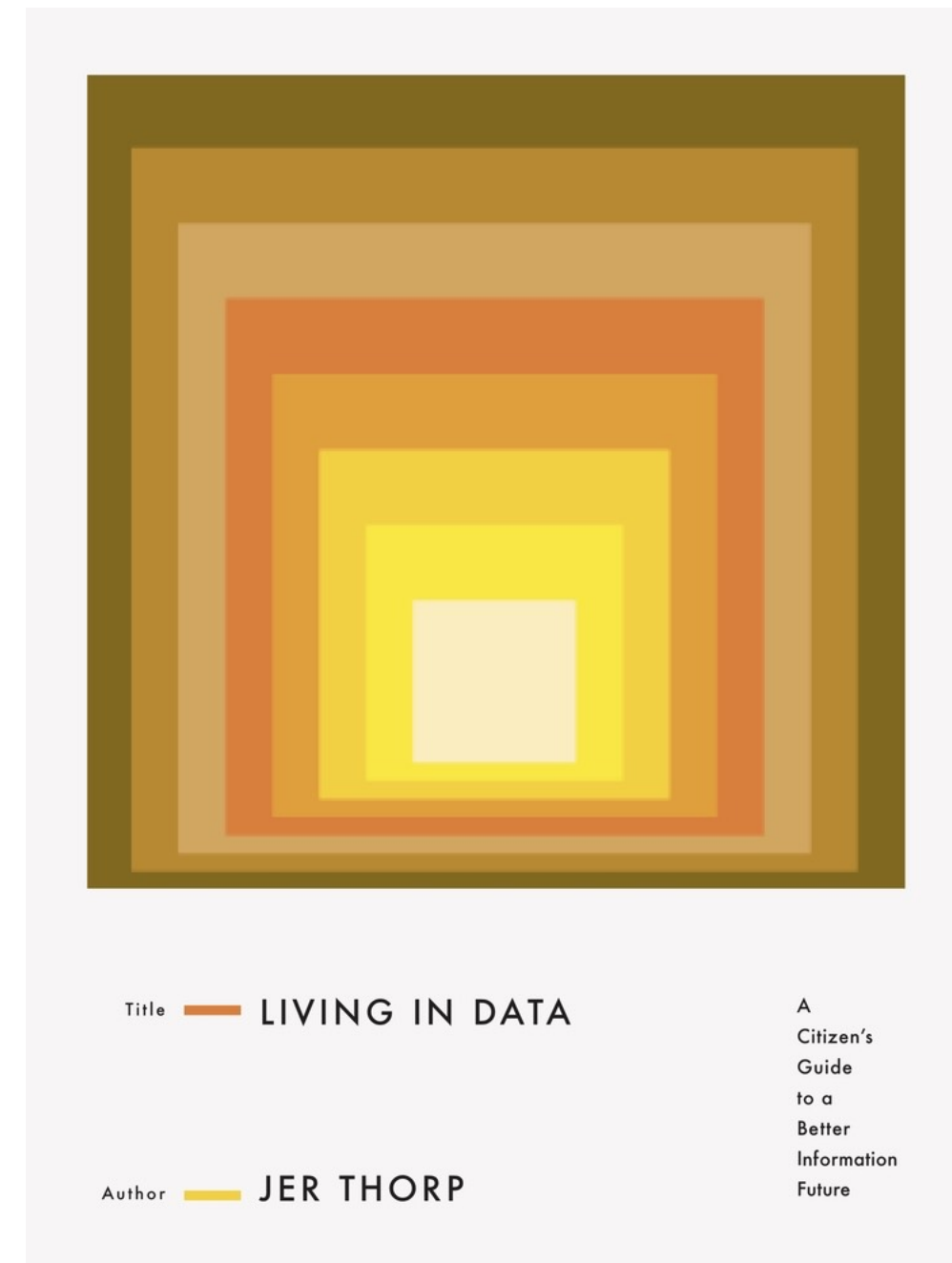
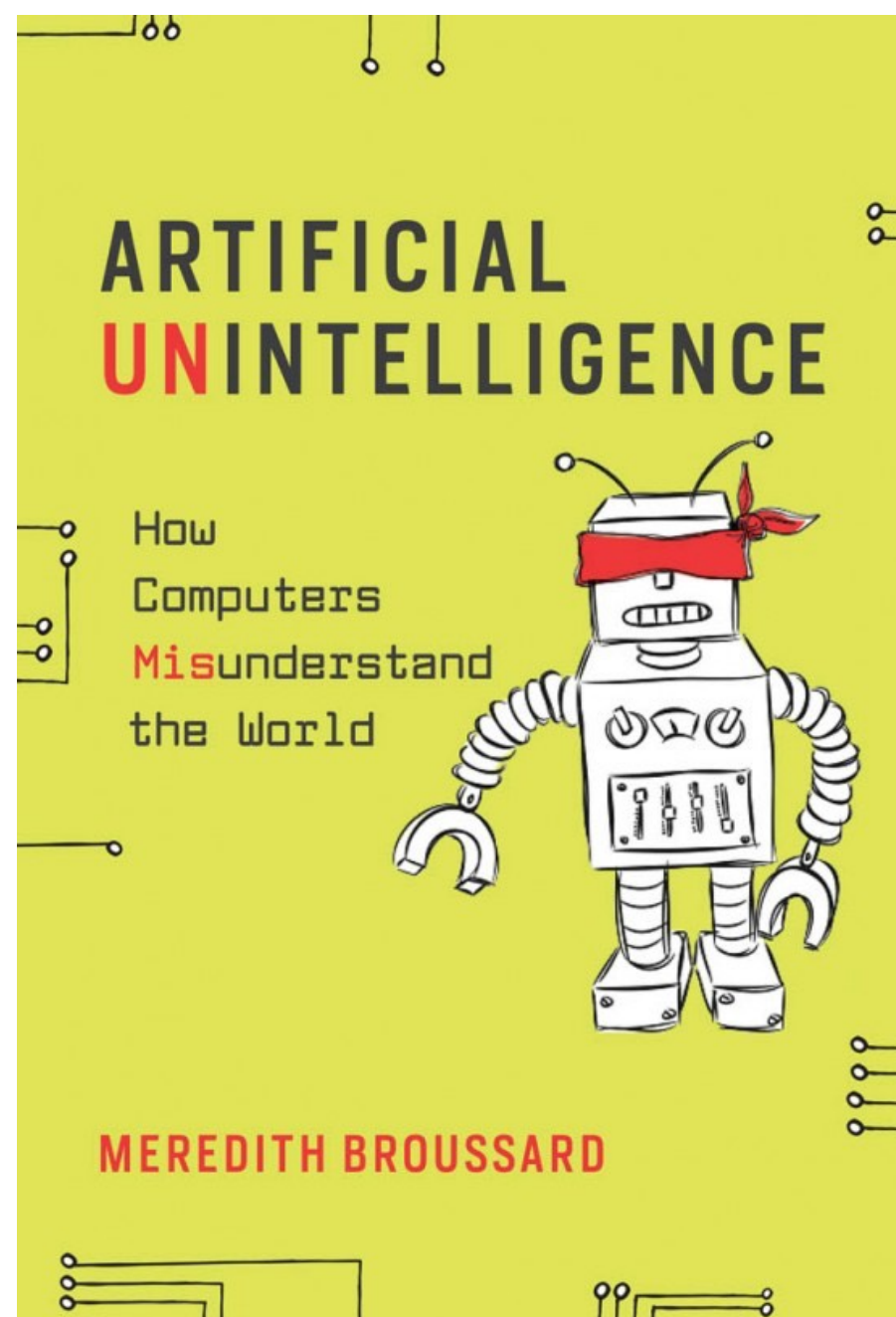
Linda Denson

Dr. Joy Buolamwini





Read/watch



STAT 336: Data Communication and Visualization

about accommodations early in the semester. Appointments can be made by calling 651-962-6315 or in person in Murray Herrick, room 110. For further information, you can locate the Disability Resources office on the web at <https://www.stthomas.edu/student-life/resources/disability/>.

Collaboration and generative AI

The goal of this course is to help you develop and strengthen your skills as a data communicator. It is acceptable to work collaboratively with other students, or use generative AI (such as ChatGPT) to help you understand topics, debug code, and clean up written work. However, I expect all work submitted to be **substantively yours**. This means that copying and pasting large blocks of code or major textual elements from another student, internet websites, or generative AI systems, is not acceptable. If I suspect work you submitted is not substantively yours, I may ask you to come in and discuss the work with me. If you cannot explain how you arrived at the solution, and restate the communication in your words, you will receive no credit. All students are bound by the [Undergraduate Student Academic Integrity Policy](#). Cases of dishonesty, plagiarism, etc., will be reported to the dean.

Resources

Course website and other technology

Canvas will be regularly updated with lecture handouts, project information, assignments, and other course resources. Reading discussion will take place on Perusall.

This course will include several pieces of industry-standard software, most crucially Tableau and R. No prior experience with any of the technologies is required for this course.

On this page

- About the Course
- Grading
- Policies and Expectations
- Common Good Community-Engaged course
- Inclusive classroom
- Stressor Statement
- Bias Reporting
- Sexual Harassment and Title IX Reporting
- Financial Hardship
- Disability Statement
- Collaboration and generative AI**
- Resources
- Tentative Schedule
- Acknowledgments

STAT 336: Data Communication and Visualization. Amelia McNamara, University of St Thomas
<https://ameliamn.github.io/STAT336/#policies>

STAT/COMP 212 - Syllabus

https://lmyint.github.io/212_spring_2024/syllabus.html#artificial-intelligence-ai-use

Artificial intelligence (AI) use

Learning to use AI tools is an emerging skill that we will explore together in this course. I expect you to use AI (ChatGPT, Google Bard)—in fact, some assignments may require it.

Please be aware of the limits of AI:

- AI does not always generate accurate output. If it gives you a number, fact, or code, assume it is wrong unless you either know the answer or can check in with another source. AI works best for topics you already understand to a sufficient extent.
- If you provide minimum effort prompts, you will get low quality results. You will need to refine your prompts in order to get good outcomes. This will take work.
- Be thoughtful about when this tool is useful. Don't use it if it isn't appropriate for the case or circumstance.
- The environmental impact of AI should not be ignored. The building and usage of AI tools consumes a lot of energy (see [here](#) and [here](#)). For this reason, we will be very thoughtful about when we use AI and will discuss other sustainability behaviors that we can incorporate into our lives to offset this usage.
- AI is a tool, but one that you need to acknowledge using. Any ideas, language, or code that is produced by AI must be cited, just like any other resource.
 - **How to cite AI:** Please include a paragraph at the end of any assignment that uses AI explaining what you used the AI for and what prompts you used to get the results. Failure to do so is in violation of the academic integrity policy at Macalester College.

If you have any questions about your use of AI tools, please contact me to discuss them.

On this page

- What is this course about?
- Course communication
- Guiding values
- How to thrive and what to expect
- Grading and feedback
- Textbooks
- Other policies
- Late work
- Academic integrity
- Artificial intelligence (AI) use**
- The environment you deserve