

Training Data Journalists In the AI Era

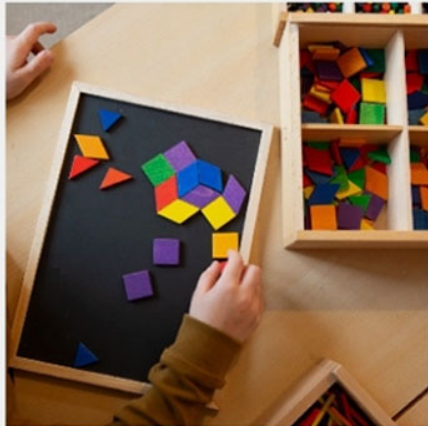
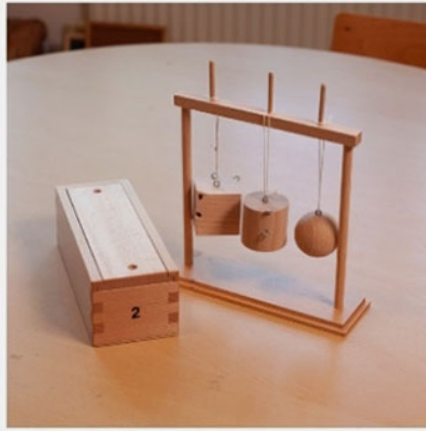
Rahul Bhargava

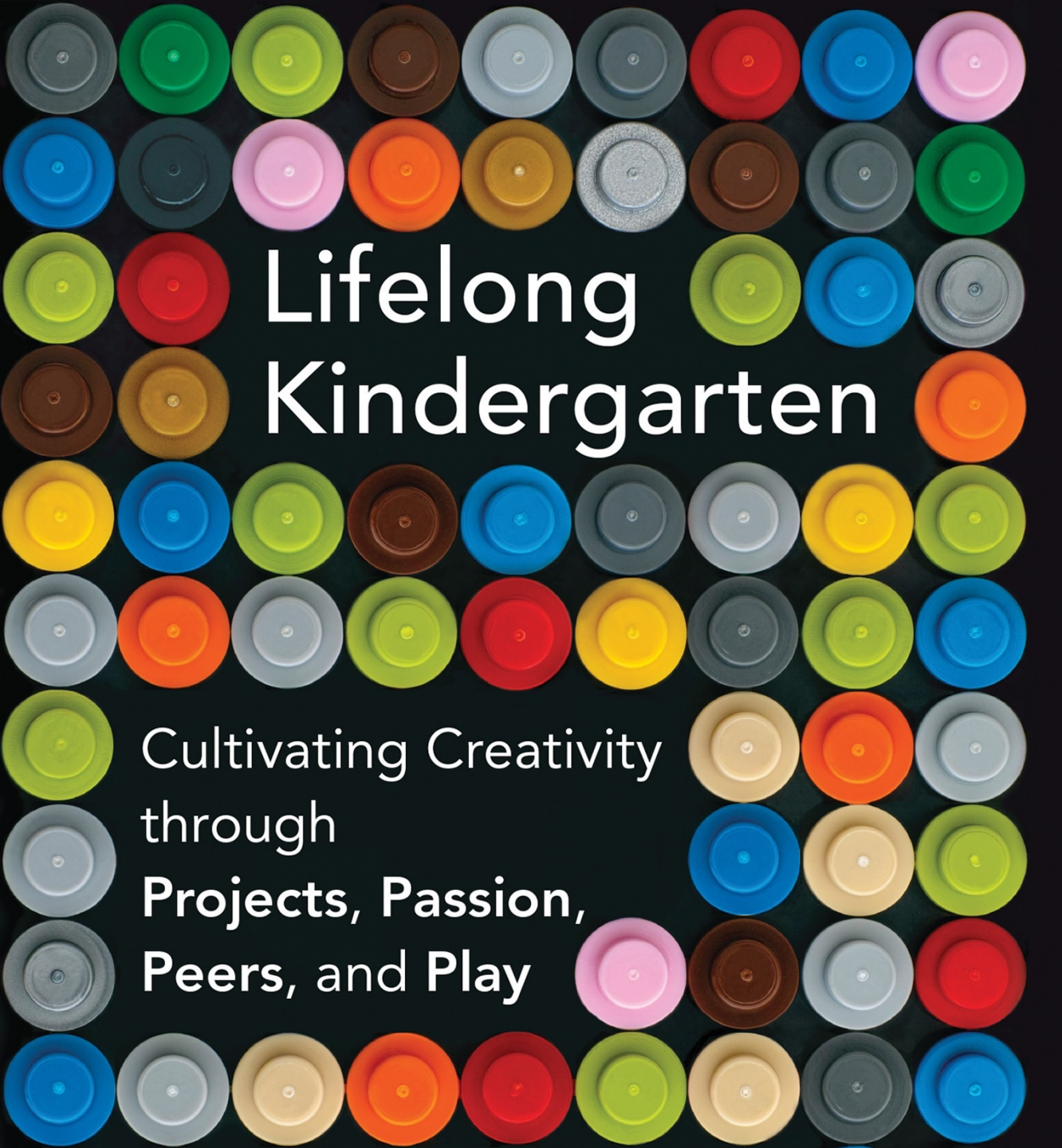
dataculture.northeastern.edu
r.bhargava@northeastern.edu
@rahulbot



DATA
CULTURE
GROUP







Lifelong Kindergarten

Cultivating Creativity
through
Projects, Passion,
Peers, and Play



Community Data

Oxford University Press

communitydatabook.com

DATA
CULTURE
GROUP



OXFORD



*creative approaches to empowering
people with information*

RAHUL BHARGAVA

Storybench

Exploring data and digital storytelling. Northeastern's School of Journalism

1 updates: how g news routines



February updates in AI and journalism: Continuing legal tensions, newsroom experiments and emerging collaboration



AI-generated image created with ChatGPT

February 6, 2025

By Rahul Bhargava And Vivica D'ouza

From lawsuits that may define AI's limits to new partnerships sparking innovation, here's a look at the month's notable developments that are reshaping the intersection of AI and journalism.

Continuing Legal Tensions

The year opened with the news that [The New York Times is taking OpenAI to court](#). According to the complaint filed by the Times, OpenAI could be on the hook for billions of dollars in damages over largely copying and using the newspaper's archives. The lawsuit also calls for the destruction of ChatGPT's dataset. NPR's Abby Alden says the lawsuit could set a crucial precedent on [whether large language models can train on copyrighted text](#).

Meanwhile, Bloomberg columnist offered varied perspectives on ["how journalism will adapt to the age of AI"](#), pointing to potential regulation and accountability. John McInerney described how an algorithm trained on satellite imagery uncovered 2,000 suspicious ship-to-ships oil transfers between Jan. 2020 and Oct. 4, 2024, in a column too long for most news sites.



December 9, 2025

By Vivica D'ouza

As 2025 winds down, AI in journalism feels less like a side project and more like the infrastructure underneath reporting, distribution and even business models. Newsrooms are building in-house systems, tinkerers are building AI experiments in newsletters and fundraising, and watchdogs are scrambling to keep up with new risks, from AI browsers turning through paywalls to scammers gaming news outlets.

This fall's AI + journalism wrap-up shows a field trying to move beyond type into durable workflows — while also confronting the mess AI is leaving in its wake.

"Hyper-cafinated interns": AI moves deeper into reporting workflows

One of the clearest themes this fall is AI taking on grunt work so humans can stay focused on accountability reporting.

At The Connecticut Mirror, AI is essentially a "hyper-cafinated intern" helping a small newsroom with its systems in the capital. [It's a look to come from elsewhere](#), because

[www.storybench.org/...in systems in the capital. It's a look to come from elsewhere](#), because



October 9, 2025

By Vivica D'ouza

Journalism's AI summer is over, and the reckoning has begun. Fast-moving trials give way to messy debates. How far should something other than what artificial intelligence actually means for the news business. Newsrooms are building chatbots. French unions are



June 18, 2025

By Rahul Bhargava, Peiyao Hu, Vivica D'ouza

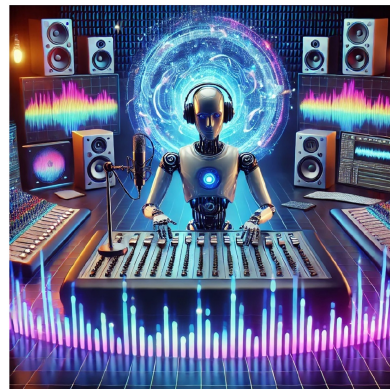
As the spring turns to summer, the intersection of AI and journalism is marked by fresh experiments and deepening debates over the future of storytelling. Here's a look at where things stand now.

What's happening in local newsrooms

Some of the most promising AI applications in journalism are emerging from small and non-Western newsrooms. In 2024, publishers are shifting from cautious trials to full integration, according to WAN-IFRA. Meanwhile, The Haystack Journalism Institute reports on how a [small-town newspaper in Illinois](#) is piloting an AI-powered content management system that's driving measurable engagement gains. These developments suggest the future of AI in journalism may not be led by big brands—but by smaller, scrappier players on the ground.

Local newsrooms are tapping AI to new heights of labor. [Statista](#) reported Gamnet, owners of USA Today and hundreds of local news outlets, are using Express, their newly introduced AI tool, to generate new content from existing content, especially for rewriting community announcements. [Other newsrooms like Cleveland and Midwest College News](#) are AI

Winter AI + journalism updates: Is AI breaking boundaries or breaking rules?



AI-generated image created with ChatGPT

December 10, 2024

By Rahul Bhargava, Peiyao Hu, Vivica D'ouza

As 2024 winds down, the relationship between AI and journalism is heating up with new lawsuits, licensing deals and AI usage in newsroom practice. Here's a summary of the latest developments.

Fall update: AI and journalism



Generated by Adobe Firefly, a generative AI program

October 24, 2024

By Rahul Bhargava, Peiyao Hu, Vivica D'ouza

While we prepare for the cold winters and shorter days, here's some need-to-know news about what's happening at the intersection of AI and journalism.

AI's role in newsrooms

Postner reports that ESPN is going ahead with having AI help cover "under-served sports" for game recaps. These AI-generated summaries will be reviewed by a human editor to ensure accuracy and quality in a release. ESPN said the move will help them "focus on their more differentiating feature, analysis, investigative, and breaking news coverage." Not everyone sees it as a win. [The Atlantic's Joe Zeff](#) says Luis Paez-Pumar wrote that the sports giant, "is leaning up its coverage of underrepresented sports... not by hiring people who can and already do write these kinds of stories, but rather by feeding existing soccer and soccer journalists' work into a machine aimed at making them obsolete."

In the Pk, Ole Zachmann, head of AI and News Strategy at Swedish Radio, discusses using the back-to-speech translation model they developed to attract 1.4 million new listeners with hearing difficulties. He also shares insights on [how they're using AI to improve content recommendations](#).

Senior Editor for AI Initiatives at The Economist's Lydia Steggle told Press Gazette that AI translation will never be a solved problem, although it has reached the point where it is good enough for the publication to have introduced [AI-powered on-air translations](#) in its budget daily briefing, [Economi](#).

Your Early Fall AI+Journalism Updates



AI-generated image created with Adobe Firefly

September 11, 2024

By Rahul Bhargava

Welcome back from the summer. So much has happened at the intersection of AI and journalism while you've been at the beach. Here's my latest summary of what caught my attention.

Skill Building

The Online News Association (ONA) has launched a huge new effort to build AI skills in newsrooms with support from various funders. Follow their [AI in Journalism Initiative](#) for trainings, reports, case studies, coverings, and more. Keep an eye on [ONA's](#) for updates.

A recently launched collaboration between the London School of Economics and Google News Initiative is offering funding to help put those skills to use. Read about the [AI-powered Journalism Project](#) and the [AI-powered Journalism Project](#) and the [AI-powered Journalism Project](#).

New Strategies for Using Artificial Intelligence in Journalism



May 16, 2024

By Rahul Bhargava

Here's the fifth in our series of round-ups on the latest in artificial intelligence and journalism. [This one's about ethics](#). This time around it feels like the rubber has really hit the road, with lots of work on strategies and examples being shared around the web. These insights are brought to you by Northeastern University data storytelling professor Rahul Bhargava.

What's Your Newsroom AI Strategy?

Larger news organizations seem to be expanding effort on conceiving artificial intelligence strategies. Just as one example, in February the [Washington Post announced](#) that [Senior Editor for AI Strategy and Innovation](#), in March [NPR announced](#) that [New York Times AI editor](#) on how the foundational principles of journalism are worth returning to, and how that plays out in concrete examples of AI-assisted reporting.

This is causing lots of reflection and strategic thinking at the C-suite level as well. The Newsroom Robots podcast [reviewed the Dutch foundation for public broadcasting](#), and [NewsLab](#) touched on AI with the [Financial Times CEO](#).

Can AI Really Help You Report?

We also found a lot more write-ups from journalist and technologist, reflecting on how AI tools





ProPublica journalists walk off the job in first U.S. newsroom strike over AI

On the picket line in New York, union leaders said they expect “more concentrated conflicts” over AI in the news industry.

By **ANDREW DECK** April 9, 2026, 1:22 p.m.



On Wednesday, roughly 150 members of the Propublica Guild, one of the largest nonprofit newsroom unions in the country, went on a 24-hour strike.

About two dozen Guild members picketed ProPublica’s headquarters in New York City’s Hudson Square neighborhood during working hours, as simultaneous picket lines formed in front of the publication’s offices in Chicago and Washington D.C. On the uncharacteristically cold April morning in Manhattan, strikers bundled up in winter gear as they chanted and carried signs reading “ProPublic Workers: Deserve Fair Pay” and “Thoughts Not Bots.”

The Guild has been negotiating its first collective bargaining agreement for two and a half years, and the one-day action was intended to put new

What’s the best way to follow how the news is changing?

Our daily email, with all the freshest future-of-journalism news.

your email address here

[Subscribe](#)

THE LATEST FROM NIEMAN LAB



Trump can’t get the media to use “Gulf of America.” Will “Lake America” be any different?

JOSHUA BENTON



Trump administration says a New York Times win in OpenAI lawsuit would “threaten national security” and hurt small newsrooms

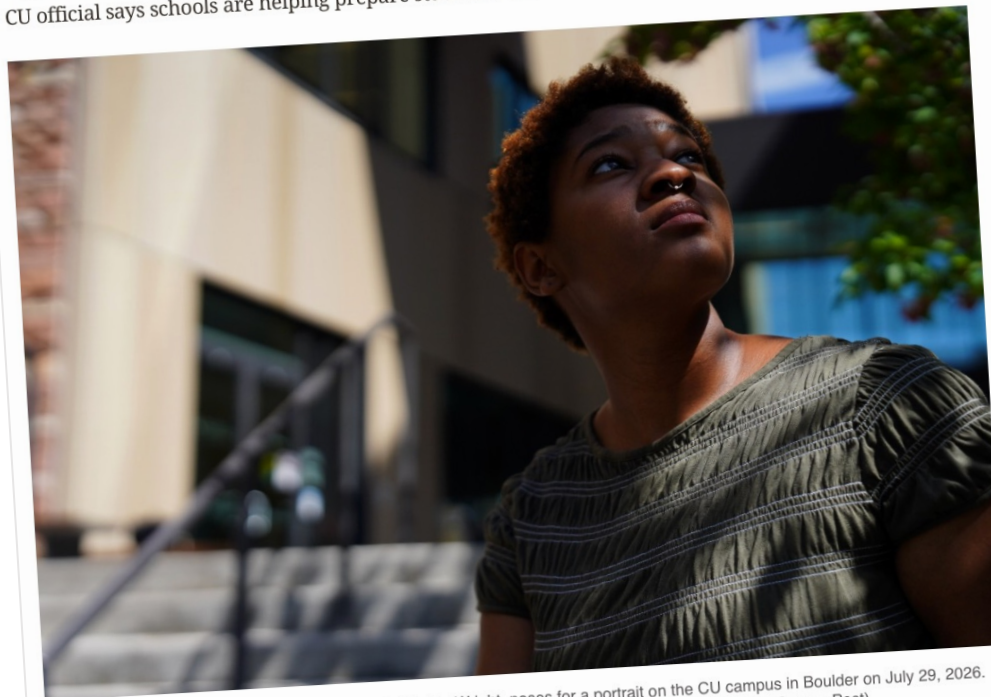
ANDREW DECK



How one newsletter writer is building a

As Colorado universities align with ChatGPT, student AI resisters take a stand against the ‘plagiarism machine’

CU official says schools are helping prepare students: “To pretend AI doesn’t exist... would be doing them a disservice”



University of Colorado Boulder graduate student, Mirakle Wright, poses for a portrait on the CU campus in Boulder on July 29, 2026. Wright is studying students who are resisting the use of generative AI. (Photo by Harmon Dobson/The Denver Post)



By **ELIZABETH HERNANDEZ** | ehernandez@denverpost.com | The Denver Post
PUBLISHED: August 7, 2026 at 4:00 AM MDT | UPDATED: August 7, 2026 at 9:33 AM MDT

Mirakle Wright thought her peers’ opposition to generative artificial intelligence in a world increasingly beset with the technology evoked the convictions held by wartime pacifists.

“People have such negative feelings about AI. They’re becoming resisters — conscientious objectors,” said Wright, a University of Colorado Boulder graduate student.

An expanding pool of research on AI’s impact on education exists, but Wright said she couldn’t find much literature about holdout college students avoiding the pull of popular generative AI



tell the truth

serve your audience

verify everything

be independent

be a watchdog

create a forum

stay relevant

make things visible

use your conscience

**OK, but
show me
your source.**

**My teaching focuses on
thoughtful use of AI,
critiquing misuse of AI,
& principled refuse of AI.**

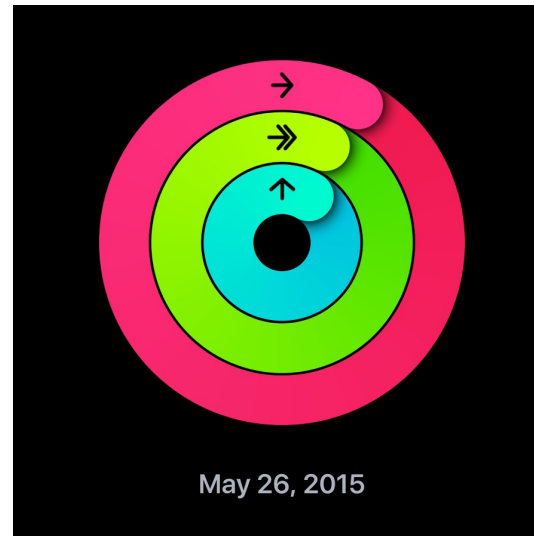
How can we learn by
being specific,
trying things out,
centering our principles
& staying critical?

Learn by being specific.

Entirely new ways to stay in touch.

Apple Watch makes all the ways you're used to communicating more convenient. And because it sits right on your wrist, it can add a physical dimension to alerts and notifications. For example, you'll feel a gentle tap with each incoming message. Apple Watch also lets you connect with your favorite people in fun, spontaneous ways — like sending a tap, a sketch, or even your heartbeat.

[Learn more about New Ways to Connect >](#)

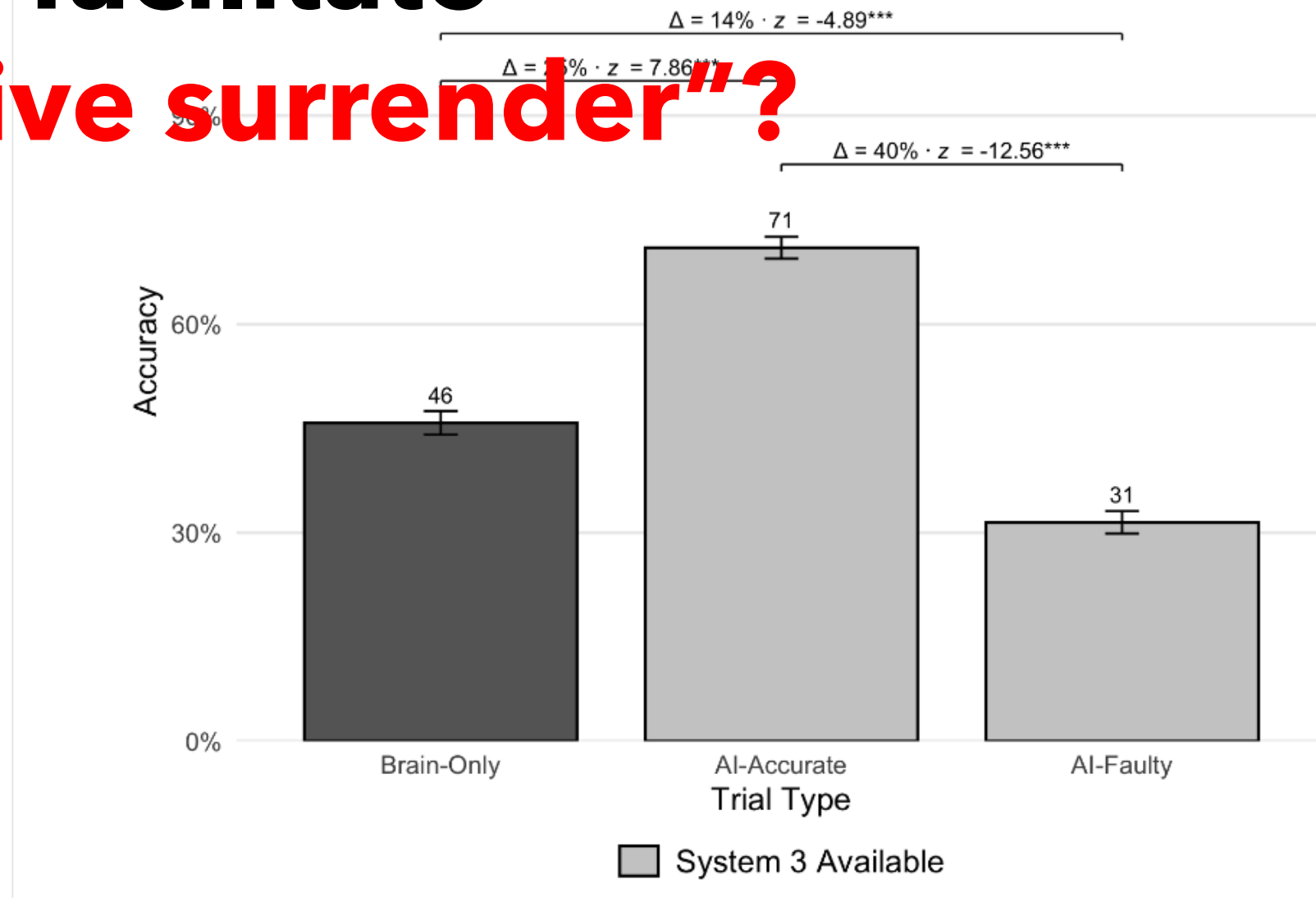


What is "AI" actually *good* at?

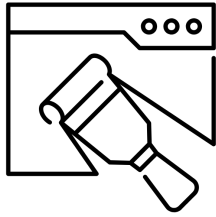
brainstorming
workflow automation
editor *coding copilot*
data scraper *chart maker*
summaries *headline generation*
translation *story discover*
writing assistant

Does AI facilitate "cognitive surrender"?

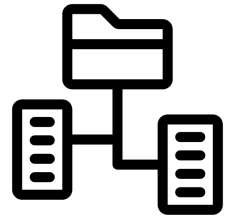
Figure 3. System 3 facilitates cognitive surrender



Be **specific** about using an AI tool for...



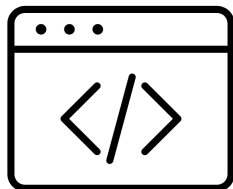
data production



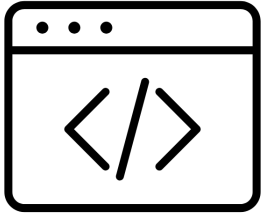
data shaping



data analysis



data representation

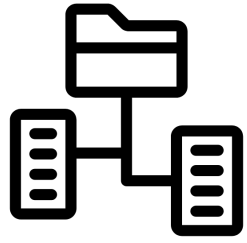


Representation: *coding interactive charts*

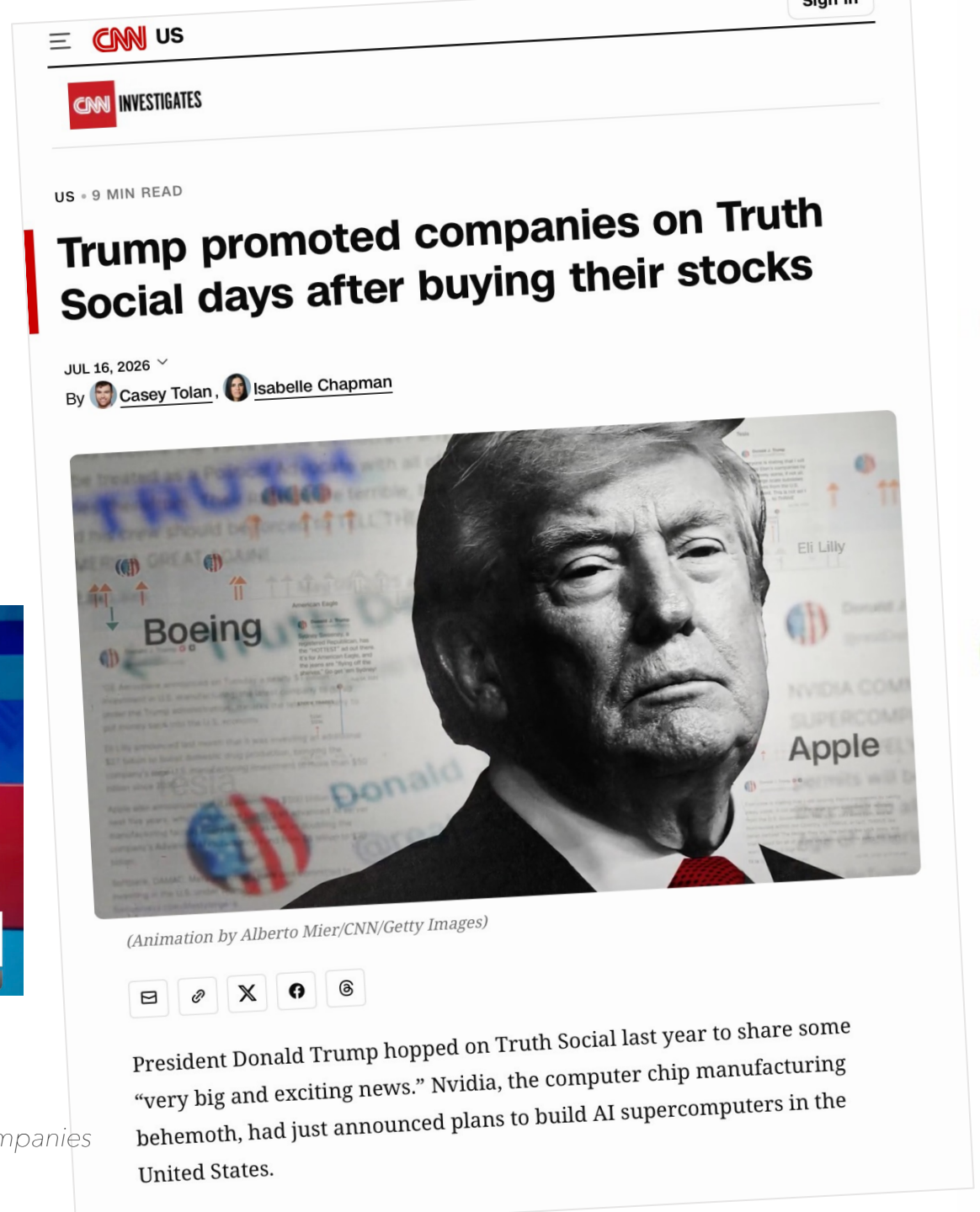


The Athletic





Shaping: *linking datasets*



**How can we learn
specific ways that AI
technologies help us?**



**Learn by
trying things out.**



You're on deadline. You have to be pragmatic.



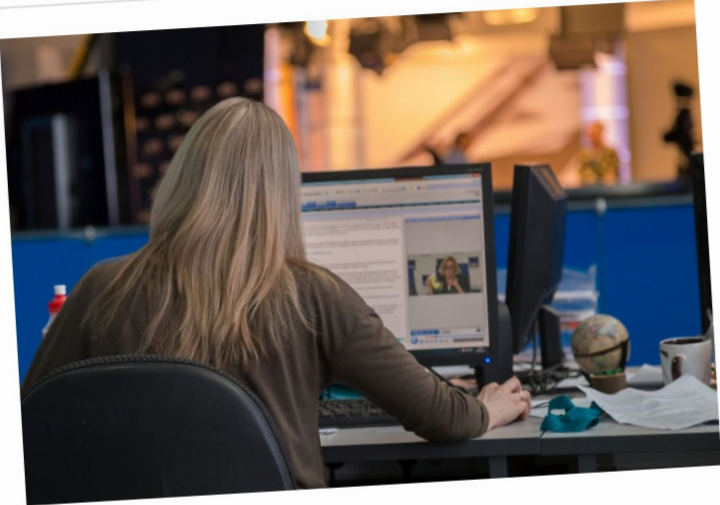
Nieman Foundation at Harvard

Nieman Fellowships Reports Lab Storyboard

NiemanLab

ABOUT SUBSCRIBE

BUSINESS MODELS | MOBILE & APPS | AUDIENCE & SOCIAL | AGGREGATION & DISCOVERY | REPORTING & PRODUCTION | TRANSLATIONS



Journalists may see AI as a threat to the industry, but they're using it anyway

Although AI use is now widespread among U.K. journalists, they still see it as much more of a threat than an opportunity.

By **NEIL THURMAN, SINA THASLER-KORDONOURI and RICHARD FLETCHER** Dec. 2, 2025, 12:20 p.m.

Although a few larger surveys of how journalists use artificial intelligence have taken place, some mostly surveyed early adopters and others didn't distinguish between current use and planned future use.

So we decided to survey a representative sample of journalists — in the U.K. We asked about their and their newsrooms' actual use of AI and how they perceived and approached it. The results were published in our recent [report](#) by the Reuters Institute for the Study of Journalism at the University of Oxford.

What's the best way to follow how the news is changing?

Our daily email, with all the freshest future-of-journalism news.

Subscribe

THE LATEST FROM NIEMAN LAB



Stars and Stripes shows how government-funded journalism can work — if it's free to report the truth

CINDY ELMORE



Trump can't get the media to use "Gulf of America." Will "Lake America" be any different?

JOSHUA BENTON



Are you *domesticating* AI?

Training programs

Career anxiety

AI marketing

*Global North/South
power dynamics*

*Small/large
newsrooms abilities*

*Big Tech funding
and philanthropy*

In-tool biases

Environmental harm

Demystifying
what is it?

Cultivating
building skills

Managing
finding limits

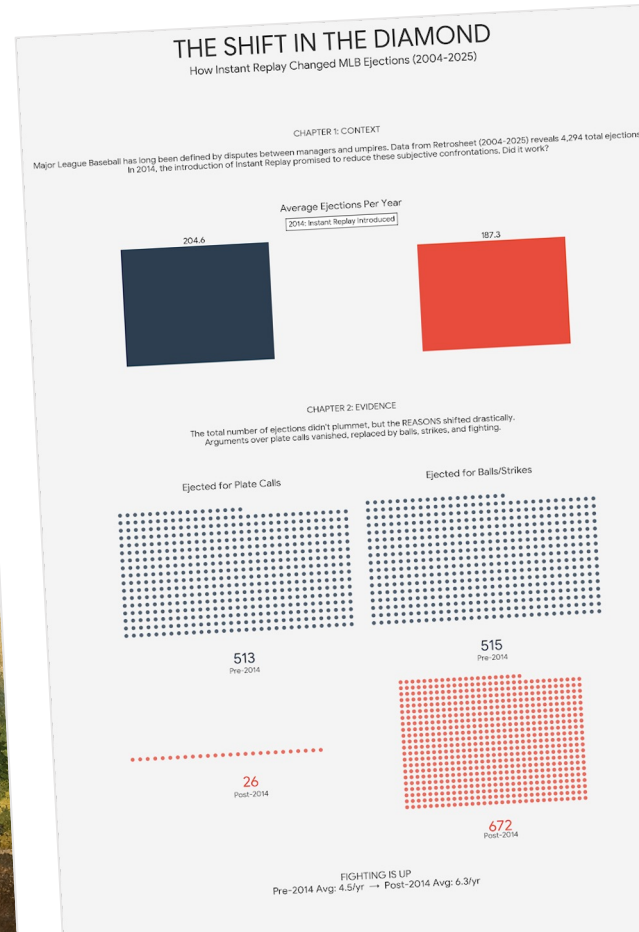
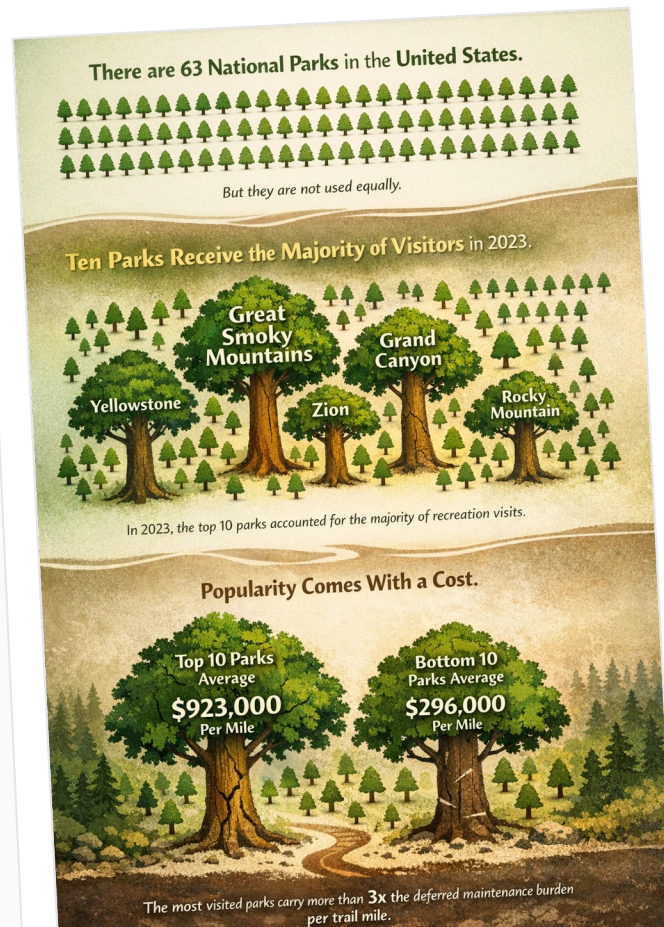
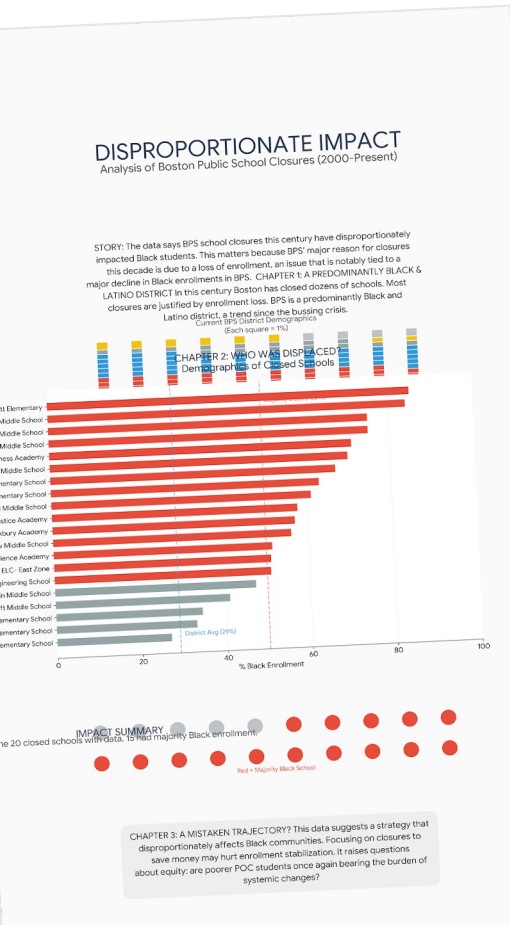
Appropriating
into practice

**Increases
confidence and
engagement
around AI**

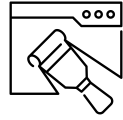
**Grows sense of
inevitability of AI
systems in the
newsroom**



From my own teaching: Generating Infographics



What are you **trying out**?



Scraping webpages



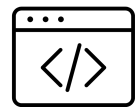
Classifying records



Parsing data out of PDFs



Finding duplicate entries



Coding interactive charts



Cross-linking unstructured data

How do engage with **harms**?

water use

energy supply costs

training data

big tech misalignment

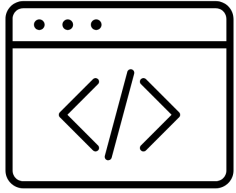
growing costs

slop-py news

out-of-date models

copyright theft

cognitive surrender



Trying to code

Unclear how much prior knowledge you need before AI assistant help you learn how to code.

However in journalism contexts:

- More of training corpus is open-source
- Technical debt less often a concern
- Local code models rapidly improving
- Domain-specific accuracy is much higher

So my worry about harms is lower.



Studying the effect of AI Code Generators on Supporting Learners in Introductory Programming

Majeed Kazemitabaar
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
majeed@dgp.toronto.edu

Barbara J. Ericson
School of Information, University of
Michigan
Ann Arbor, Michigan, USA
barbarer@umich.edu

Justin Chow
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
justinchow@dgp.toronto.edu

David Weintrop
College of Education, University of
Maryland
College Park, Maryland, USA
weintrop@umd.edu

Carl Ka To Ma
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
cma@dgp.toronto.edu

Tovi Grossman
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
tovi@dgp.toronto.edu

Generative AI in Computer Science Education: Accelerating Python Learning with ChatGPT

Ian McCulloh
Johns Hopkins University
3400 N Charles St
Baltimore, MD 21218
lmccull4@jhu.edu

Manu Gupta
Great Learning
manu.g@mygreatlearning.com

Pedro Rodriguez
Johns Hopkins University
3400 N Charles St
Baltimore, MD 21218
prodrig4@jh.edu

Viplove Raj Sharma
Great Learning
viplove@mygreatlearning.com

Anthony N. Johnson
Johns Hopkins University
ajohn260@jhu.edu

Srivaths Kumar
Great Learning
India
srivaths.kumar@greatlearning.in

Benjamin Johnson
Johns Hopkins University
bjohn159@jhu.edu

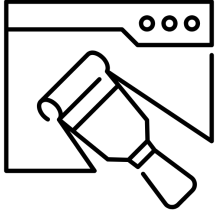
ABSTRACT

The increasing demand for digital literacy and artificial intelligence (AI) fluency in the workforce has highlighted the need for scalable, efficient programming instruction. This study evaluates the effectiveness of integrating generative AI, specifically OpenAI's ChatGPT, into a self-paced Python programming module embedded within a sixteen-week professional training course on applied generative AI. A total of 86 adult learners with varying levels of programming experience completed asynchronous Python levels of programming experience completed asynchronous Python instruction in Weeks three and four, using ChatGPT to generate, interpret, and debug code. Python proficiency and general coding knowledge was assessed across 30 different assessments during the first 13 weeks of the course through timed, code-based evaluations. A mixed-design ANOVA revealed that learners without prior programming experience scored significantly lower than their peers on early assessments ($\eta^2 = .10-.17$). However, following the first 13 weeks of the course through the instruction module, these

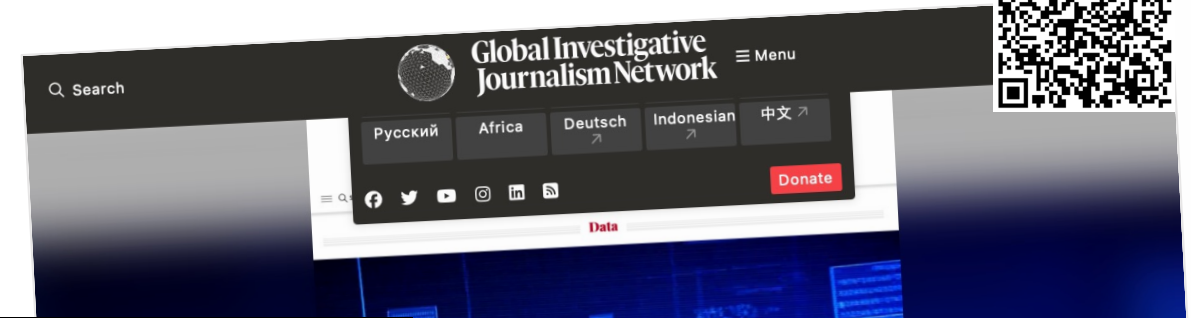
1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) technology and the growing demand for digital skills across industries has intensified the global need for reskilling and upskilling in the workforce. Among the most critical competencies in the modern workforce is programming literacy, particularly in widely used languages such as Python. Python's versatility and readability have made it the language of choice not only for software development but also for data science, automation, and AI applications. Independent market research contracted by Johns Hopkins University reveals that 35% of software-related job postings require Python proficiency and the year-on-year projected growth in Python proficiency is +24%. Yet, traditional approaches to teaching programming—centered on lectures, rote memorization, and repetitive coding practice—often require extended instructional timelines and produce uneven learning outcomes, particularly for adult learners or professionals transitioning into technical roles.





Data Production: *data logging*



The days start out manageable, as **Sathya, the tiffin cook**, wakes up to a quiet street in Parry's Corner even before the sun rises, in the last hours before the discomfort begins.

At around 9:00 a.m., she is ready with her meat and fish prepared, spices ground, and vegetables chopped.

SHARE

The Hindu Is Embedding AI Into Its Journalism

April 10, 2026

READ THIS ARTICLE IN

বাংলা English

中文

Months, journalists at The Hindu parsed nearly 22 million voter records across states, built an election results interface without writing a line of code and assembled low-cost heat sensors to measure how different workers experience extreme temperatures.

At work, Srinivasan Ramani says, was by large language models (LLMs), used to generate prose, but to process documents, and structure investigations.

Deputy national editor and senior associate at The Hindu, outlined this approach at our [AI Forum Bangalore](#) 2026.

It was not on automation of writing, but on how low-code and AI-assisted tools are reshaping data journalism workflows inside a legacy newsroom.

He said, "is a very sophisticated intern. You tell it exactly what to do. It does it. But it's in control."

"The hypothesis was ours. The political and social context was ours. AI helped us process the scale." — Srinivasan Ramani, deputy national editor, The Hindu

RELATED STORIES



Developing a Data State Of Mind: Key Tips for Editors



My Favorite Tools: Indian Journalist and Data Designer Gurman Bhatia



What I've Learned: Lessons and Advice from The Caravan's Vinod K. Jose



How a WhatsApp Community Is Bringing India's Data Storytellers Together

[More Stories →](#)



[Explore the Resource Center →](#)

Parsing 22 Million Voter Records

One of the most extensive projects Ramani described examined India's Special Intensive Revision (SIR), a periodic update of voter rolls conducted by the Election Commission. In the latest round, authorities released records listing deleted voters and the reasons cited.

The data was not analysis-ready. It came in the form of image-based PDFs, effectively photographs of forms, in Hindi.

In Bihar alone, the team processed around 90,000 files covering 65 lakh (6.5 million) records. Tamil Nadu involved roughly 78,000 files and 97 lakh (9.7 million) records; West Bengal, about 80,000 files and 58 lakh (5.8 million) records. In total, the three

DATA
CULTURE
GROUP



Tolan, C., & Chapman, I. (2026, July 16). Trump promoted companies on Truth Social days after buying their stock. CNN.



Trying to investigate

Winners of the Agentic AI Investigative Challenge. Building systems that analyzed data to:

- Generate leads including evidence bases
- Extract metadata and create traceable citation records
- Evaluate outputs of generated code that cleans and matches

The all focused on provenance, hashing, and multiple passes at computation.



Generative AI in the Newsroom

Exploring the responsible use of generative AI in news production.

1.7K followers · 2 editors

Announcing the Winners of the Agentic AI Investigative Challenge



Nick Hagar

Follow

4 min read · Aug 27, 2026



13



1



Over 60 teams signed up for the inaugural Agentic AI Investigative Challenge, which asked them to scour Congressional lobbying data using Claude Code. The results give us a glimpse of what an AI-powered future of journalism might look like. With help from an outside group of journalists and researchers, we reviewed 200 Agent Skills submitted by our participants and selected three winners from a very competitive pool of submissions. Each of these winners offers a new perspective on how to use AI agents in a more reliable, more scalable and cost-



**How can we learn by
trying things out with AI
to see how it can help our
work?**

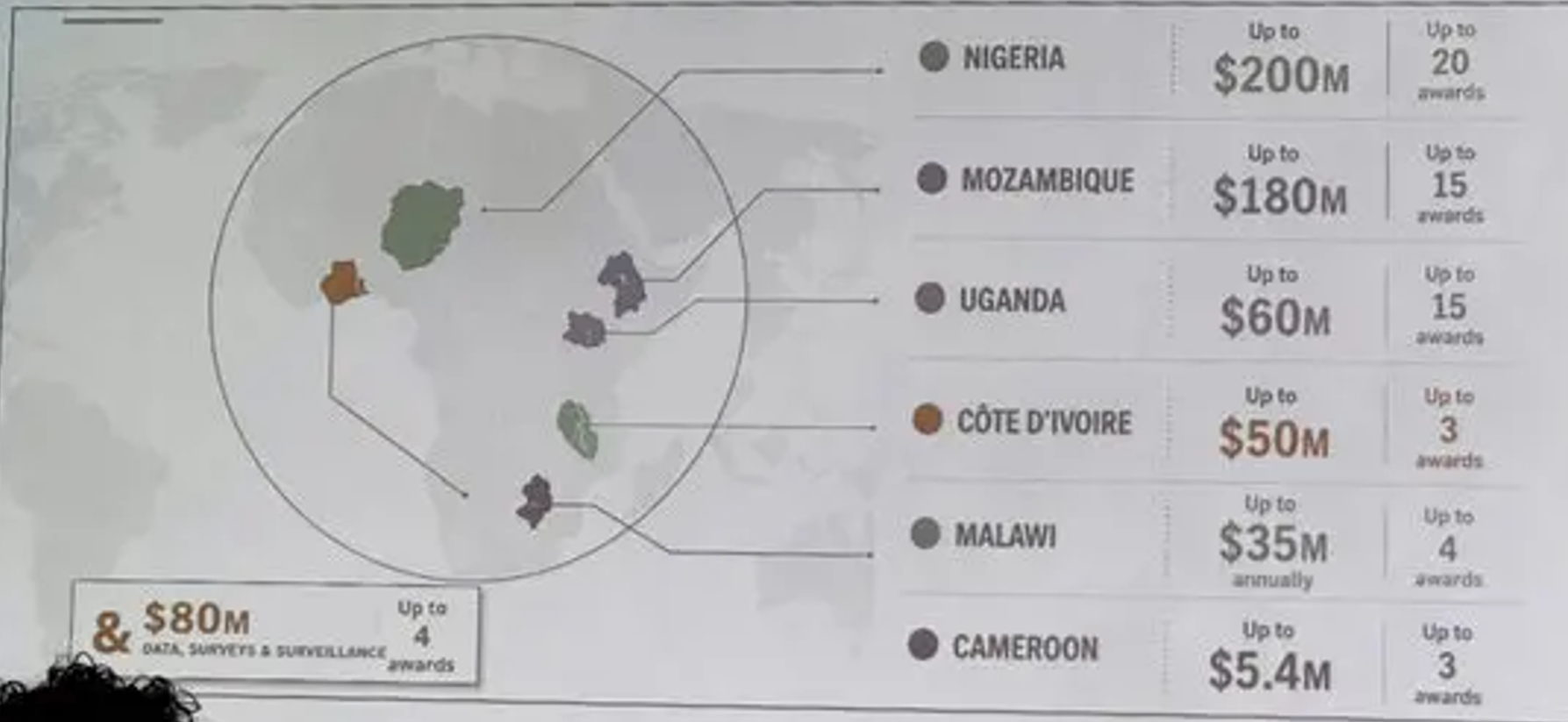


Learn by centering our principles.

- **Accountability:** Our journalists are responsible for everything we publish. If AI is used in our work, it must be reviewed by staff and meet the same editorial standards as the rest of our journalism.
- **Verification:** We treat AI-generated material as unverified source material. Because these tools can produce errors, fabricate information or reflect bias, we independently verify facts and claims before publication.
- **Transparency:** If AI plays a meaningful role in the reporting of a story, we will clearly disclose its use to our audience.
- **Integrity:** We do not use AI to generate or manipulate images, audio or video in ways that could mislead our audience. Moreover, AI tools do not replace the original reporting, analysis and judgment of our journalists.
- **Privacy and security:** We protect our sources, subjects and data by ensuring sensitive or confidential information is handled only through secure, trusted systems. We take care not to input confidential source documents into public AI tools that could compromise data privacy.



Current APS Funding Opportunities



Note: This is the second round of awards



Editors' Note: May 1, 2026

This article was updated after The Times learned that a remark attributed to Pierre Poilievre, the Conservative leader, was in fact an A.I.-generated summary of his views about Canadian politics that A.I. rendered as a quotation. The reporter should have checked the accuracy of what the A.I. tool returned. The article now accurately quotes from a speech delivered by Mr. Poilievre in April. He said, "My personal opinion is that when a member of Parliament goes back on the word they made to their constituents and switches parties, constituents should be able to petition to throw them out and have a byelection. That would put the people back in charge of our democracy rather than having dirty backroom Liberal deals by Mark Carney determine our destiny." He did not refer to politicians who changed allegiances as turncoats in that speech.



Canada speaking on Saturday at a Liberal Party convention

in Montreal. Ian Austen/The New York Times

The rising star in global centrist politics has secured a majority in the Canadian Parliament. Critics are crying foul.

Listen · 9:42 min

Share full article



282



By **Matina Stevis-Gridneff**

Reporting from Toronto

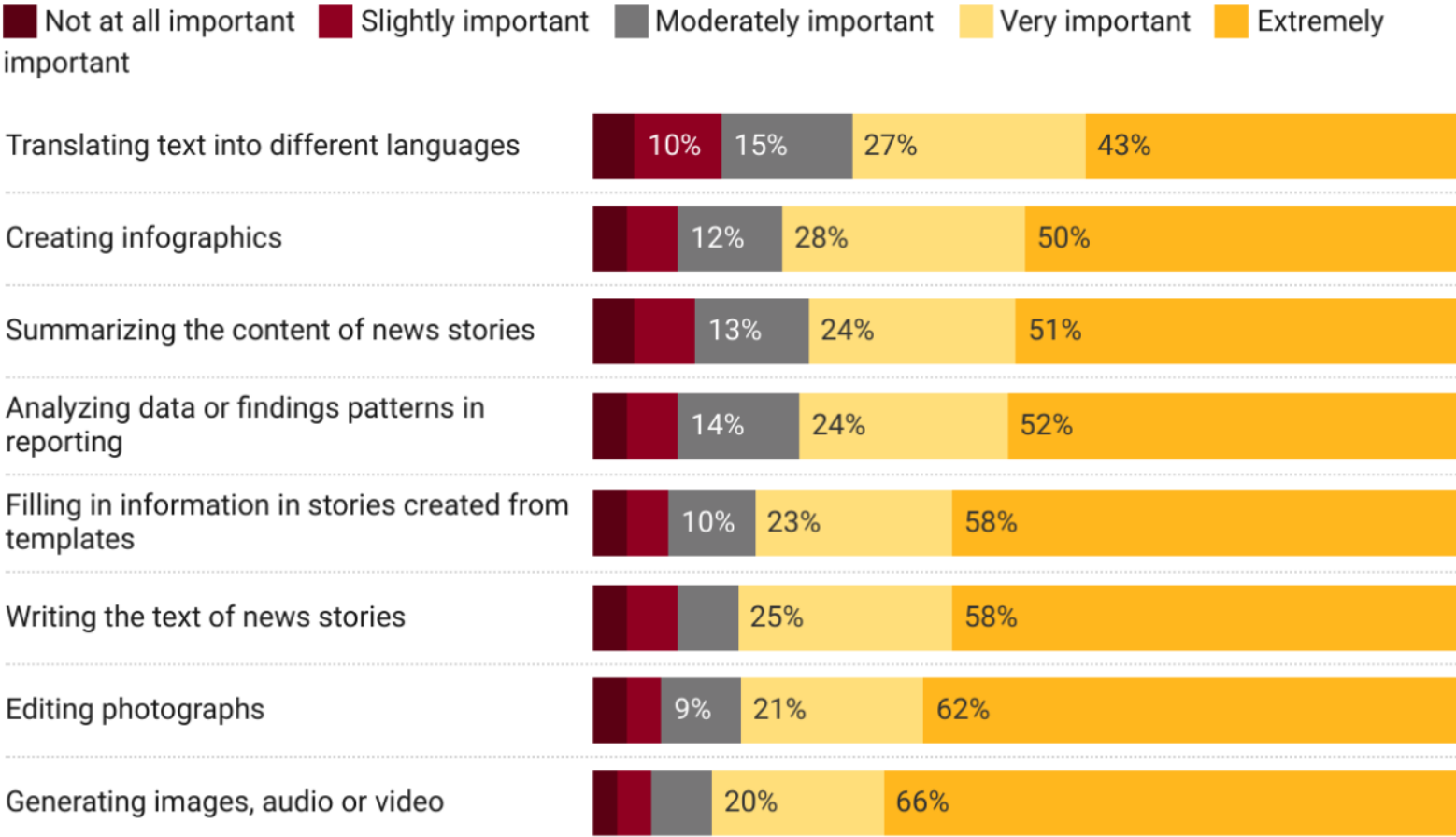
Published April 14, 2026 Updated May 1, 2026

It's Mark Carney's Canada now.

One year, almost to the day, since the stunning electoral win that made



Importance of AI disclosure varied by use case in survey of cohort audiences



Note: Respondents were asked "How important, or not important, is it for news organizations to disclose or label using AI when doing any of the following?"



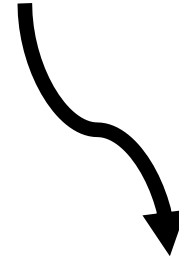
Values-Based AI Policies

What are AI tools
being used for
our students?



What rules do we
need for the
tools?

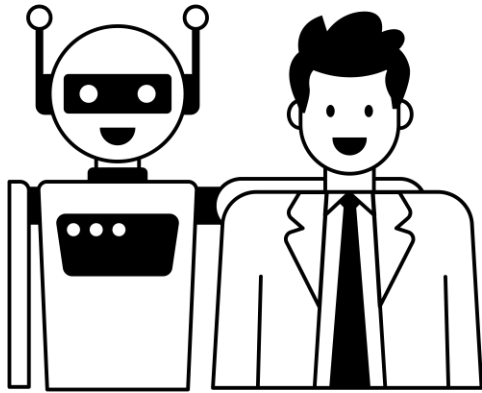
What do we care
about most?



What's an AI policy
that reinforces
those core values?

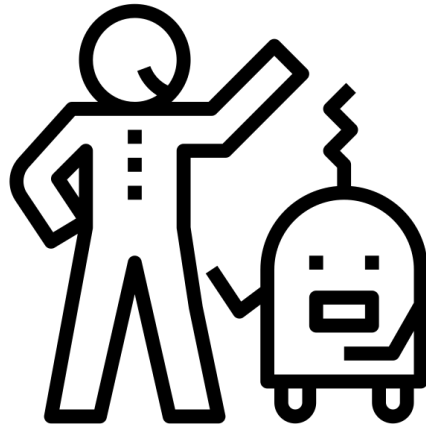


What role is AI playing in learning?



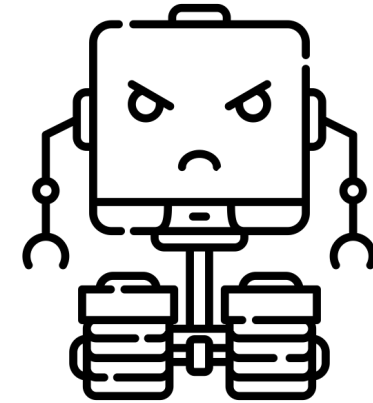
Colleague

AI is your partner, you become a "meta-gatekeeper"



Assistant

AI does grunt work so you can then use outputs



Outsider

AI is threat to reliability, you focus on critical explanation



Menü

DIE ZEIT

Abo testen

Z+ NSDAP-Mitgliederkartei

Recherchieren Sie hier die NSDAP-Vergangenheit Ihrer Familie

Die ZEIT hat die Mitgliederkartei der NSDAP aufbereitet. Mit unserem Tool lassen sich Millionen Dokumente einsehen. Suchen Sie nach Ihrem Familiennamen.

Von Gregor Aisch, Jakob Bauer, Andreas Loos, Selina Rudolph, Nico Scordialo, Shairal Sharma, Christian Staas, Julian Stahnke, Sven Stockrahm und Julius Tröger

Aktualisiert am 7. April 2026, 13:02 Uhr ⓘ

1.442 Versuche Zusammenfassen

Neu: Durchsuchen Sie jetzt auch die Personalakten der SA und SS nach Ihren Familienangehörigen

Waren Ihr Vater, Ihre Großmutter oder Ihr Ur-Opä in der NSDAP? Bis vor Kurzem ließ sich das nur über eine Anfrage beim Bundesarchiv klären

oder aktuell auf der Seite des US-Nationalarchivs, das seine



**How can we learn to
center our principles in
any scenario where we
decide to use AI tools?**



**Learn by
staying critical.**

Can we create alternatives?

Thomson: Continual Learning of Frontier Models for SovereignAI

Shengzhuang Chen*, Jerrod Parker*, Yejin Bang*, Andrew M. Bean*, Nabeel Seedat*, Stefan Winzeck†, Daniil Glazko†, Jannik Zraggen†, Fangyi Yu†, Scott Arnott†, Dietrich Trautmann†, Luca Ciuffreda†, Guglielmo Bonifazi†, Davide Romano†, Bradley Bell†, Kirsty Fielding†, Daniele Giofrè†, Tom Zielund†, Ipshita Chatterjee†, Sneha Murthy Ghantasala†, Manpreet Nanreh†, John Scoville†, Maciej Sakowicz†, Wassim Se, Jonathan Richard Schwarz†

*Primary Authors, †Core Contributors, ‡Contributors
Correspondence: First.Last@thomsonreuters.com,

ETH zürich

ETH Zurich homepage

Menu

Abstract

The development of frontier models is cornered by a number of heavily funded players, creating a diverse user base of model developers and calling for SovereignAI (an organisation's use), but often providing little concrete advice on the diversity of funding settings.

In this report, we argue that frontier performance can be improved through Continual Learning on readily available data. In our model, our Continual Learning approach tailors the training stack while introducing safeguards at the stage and seeking to make the minimal impact on the model. This strategy results in model improvements across successive model generations. Crucially, the training budgets are substantially lower than commonly used budgets (model, tool infrastructure, values & stack).

To demonstrate this, we introduce Thomson, an enhanced focus on high-stakes professional productivity improvements through data-centricity, and efficiency, we demonstrate frontier models on a wide range of domains: legal, tax & multilingualism, to comprehend a distinctive π -shaped pattern: distinct in those not explicitly targeted), while almost narrow domain adaptation.

In partnership with:  Imperial College London
Open-weight model: Thomson-1.0-5m

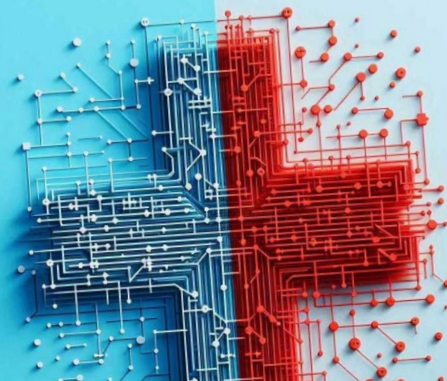
ARTIFICIAL INTELLIGENCE · INNOVATION & INDUSTRY

A language model built for the public good

ETH Zurich and EPFL will release a large language model (LLM) developed on public infrastructure. Trained on the "Alps" supercomputer at the Swiss Supercomputing Centre (CSCS), the new LLM marks a milestone in open AI and multilingual excellence.

09.07.2025 by Florian Meyer, Corporate Communications and Mélissa Anchisi, Head of AI Communications

Read · Share



tech eu News Topics Insights Funding Explorer

ARTIFICIAL INTELLIGENCE

How Danish transcription platform Good Tape grew from a newsroom hack to 2.5M users globally

In a race dominated by data harvesting, Good Tape takes the slow road — hosting its own language model, keeping data private, and earning the trust of millions of users along the way.

Cate Lawrence
14 April 2025

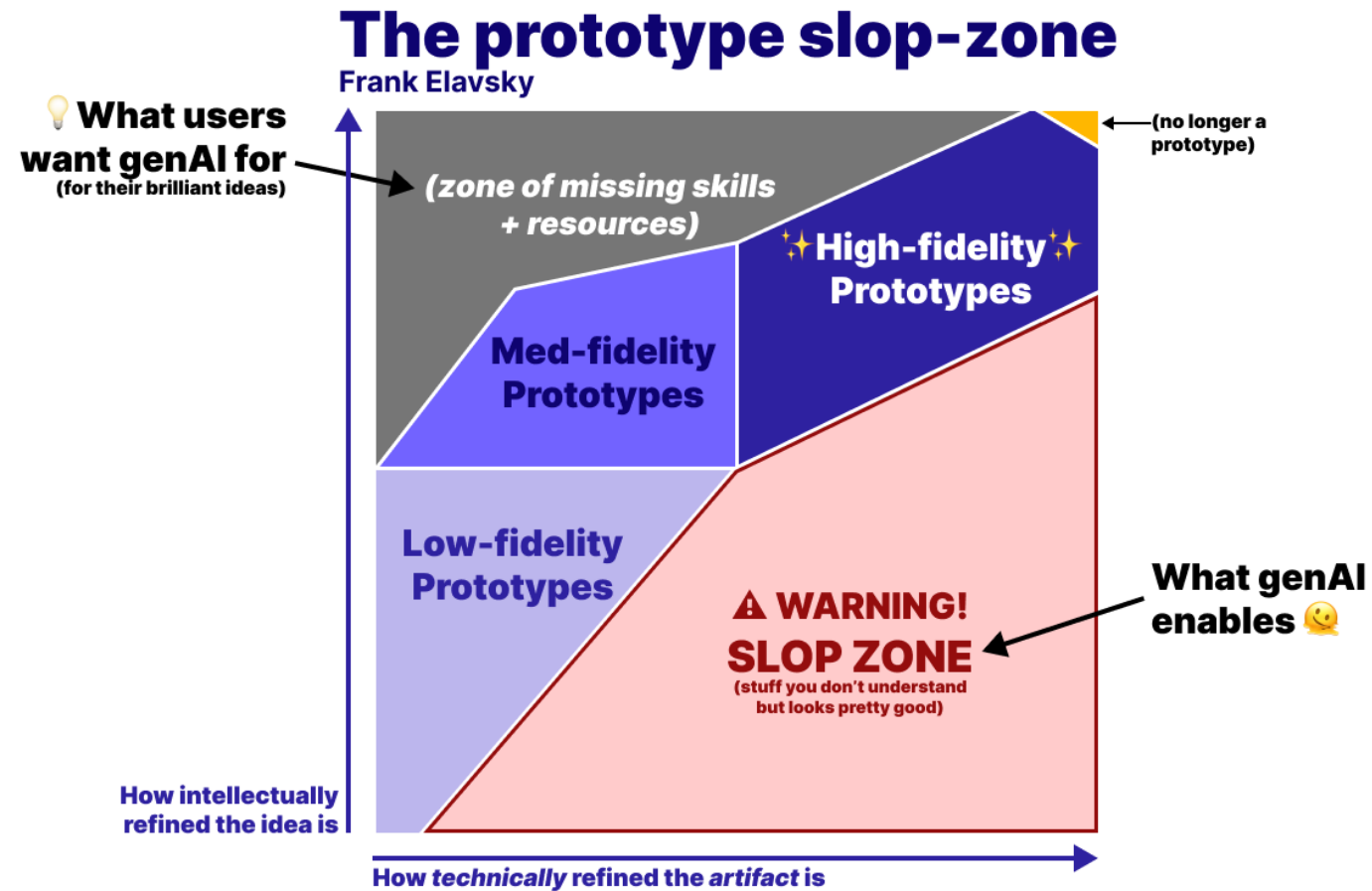


† See Acknowledgements for additional contributions

What of value is being **lost**?



Is prototyping **faster** good for learning?

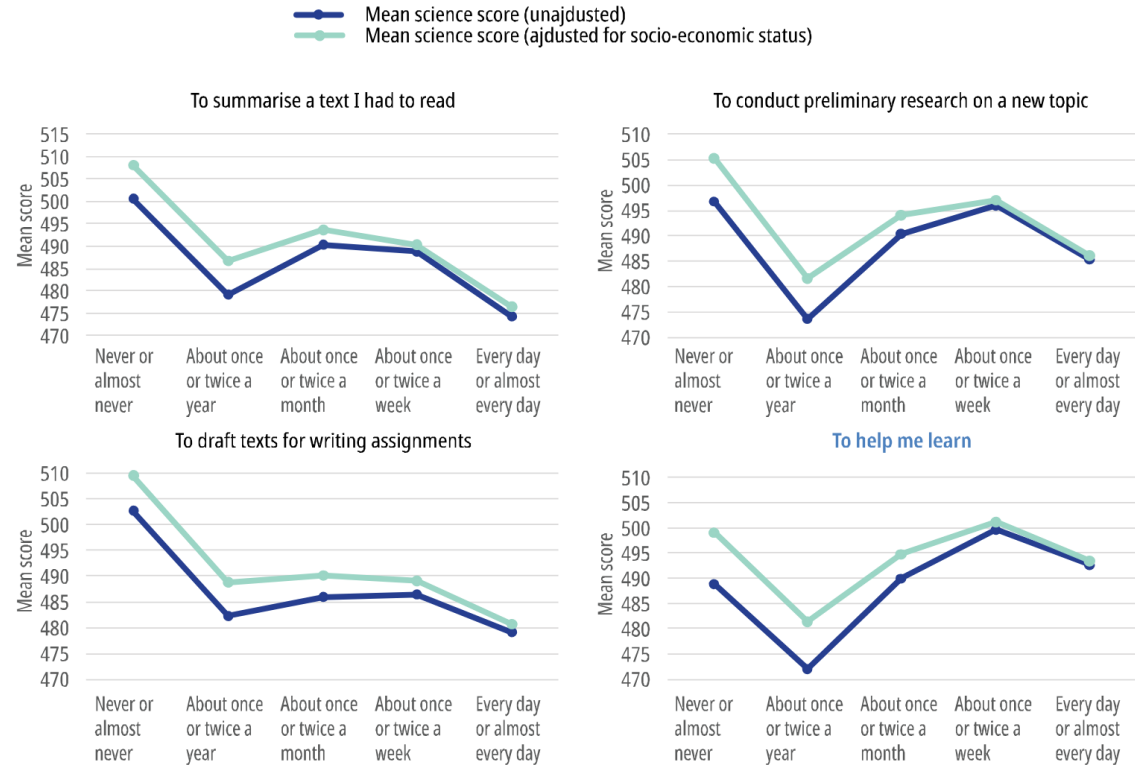


Elavsky, F. (2026, March 9). On genAI: Was prototyping really a bottleneck?

Does AI help traditional learning?

Figure I.4.13. Mean score in science, by frequency of AI use for schoolwork

Based on students' reports; OECD average



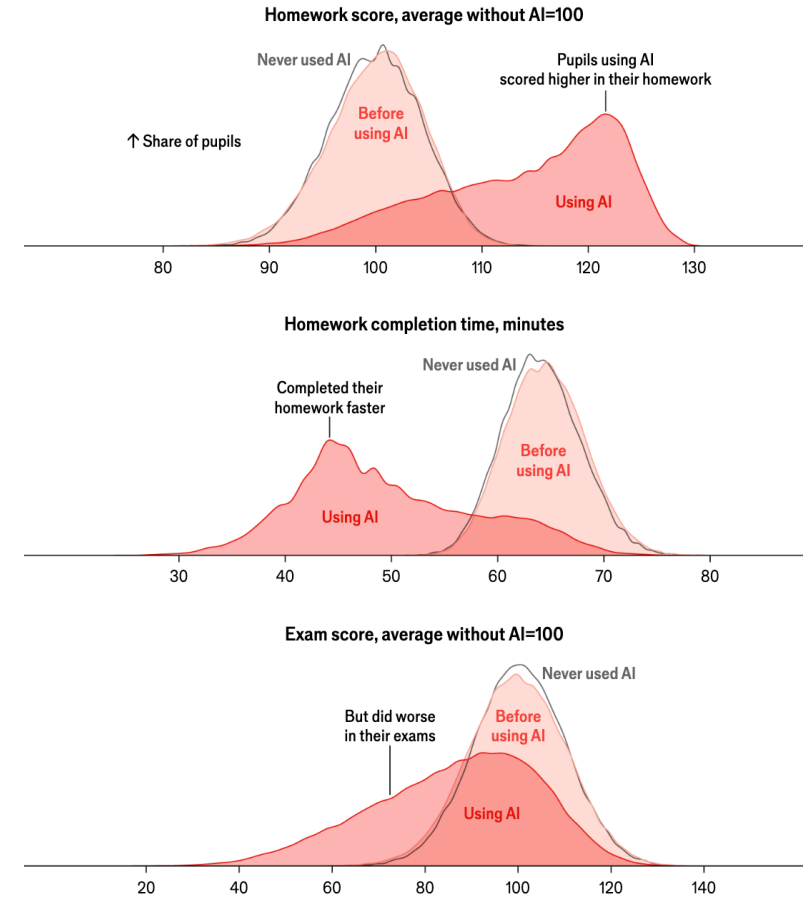
Note: Mean scores adjusted for socio-economic status are predicted science scores obtained from separate regressions of science performance on the frequency of each type of AI use, after accounting for students' socio-economic status.

Source: OECD, PISA 2025 Database, Table I.B1.4.61. See <https://stat.link/o14k0> for the underlying data.

AI ate my homework

China*, effect of using AI, January 2023-June 2025

The Economist



OECD. (2026). PISA 2025 Results (Volume I): Future-Ready Students. OECD Publishing.

Stromberg, D., Lei, V., & Wu, Y. (2026). The Generative AI Learning Penalty: Evidence from Chinese Secondary Education. SSRN.



How can we **stay
critical of the AI
tools we learn with?**

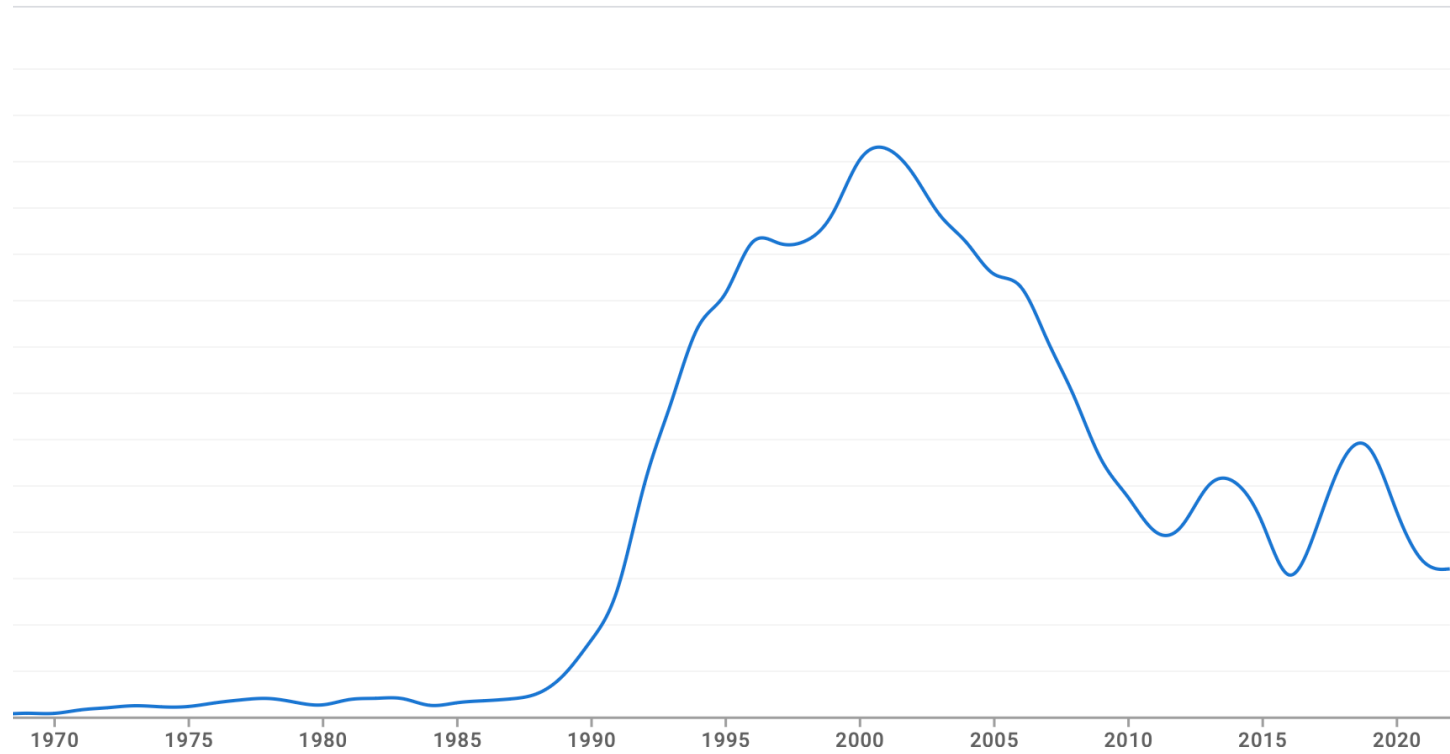


How can we learn by
being specific,
trying things out,
centering our principles
& staying critical?



SciCAR

Where science meets
computer-assisted reporting



Google n-gram viewer for "computer-assisted-reporting"

stay critical

be specific

Learn CARefully

try things out

center our principles

We learn CAREfully through:
thoughtful use of AI,
critiquing misuse of AI,
& principled refuse of AI.

AN INVITATION

C+J 2026 Oct 2-3 Northwestern



DATA
CULTURE
GROUP

A screenshot of the Northwestern University website for the Computation + Journalism 2026 Symposium. The page features a purple header with the Northwestern logo, a search bar, and navigation links for Call For Participation, Registration, Schedule, Hotels & Travel, and Organizing. The main banner shows a scenic view of Northwestern University's campus with a lake and buildings, overlaid with the text "Evanston, October 2nd-3rd 2026" and "Computation + Journalism Symposium". Below the banner, a section titled "News Beyond the Newsroom" provides a description of the symposium as a space for those interested in the intersection of journalism, news, and technology. It lists the dates as October 2-3, 2026, and the location as Evanston, IL. The page also features portraits of the opening and closing keynote speakers, Arvind Narayanan and Casey Fiesler, along with a photo of Casey Fiesler in conversation with Gina Chua. At the bottom, there is a call to action to register now, highlighting an early registration discount that ends on September 8th, and a note about discounted hotel rooms available until September 1st.

Training Data Journalists In the AI Era

Rahul Bhargava

dataculture.northeastern.edu
r.bhargava@northeastern.edu
@rahulbot



DATA
CULTURE
GROUP

