

Are We Data? (Do We Matter?)

PARTICIPATION THROUGH THE EYES OF OTHER ACTORS

Brian Ballsun-Stanton

INTS1302 Navigating Digital Worlds, Macquarie University

2026-08-18

We Want Information

COERCION AND THE DEMAND FOR INFORMATION

- Contemporary platforms obtain information through ordinary participation.



Displayed clip: [The 3:02 opening titles of *The Prisoner* \(1967\)](#)
([*The Prisoner* \(1967\), 2024](#))

People Construct Social Spaces

NETWORKED PUBLICS AS TECHNOLOGY AND SHARED PRACTICE

Publics provide a space and a community for people to gather, connect, and help construct society as we understand it.

([boyd, 2014, p. 9](#))

- Physical publics include parks, malls and sporting events.
- Hybrid publics combine co-present activity with networked communication.
- Online publics range from small forums to transnational movements.
- A networked public is both a technological space and an imagined community.

Four Affordances Shape Participation

WHAT CAN BE REMEMBERED, SEEN, SHARED, AND FOUND

- Affordances connect a space's design with people's practices and expectations.
- **Persistence** concerns how long content can remain available.
- **Visibility** concerns who can encounter the content.
- **Spreadability** concerns how easily content can be copied and shared.
- **Searchability** concerns how easily content can be found.
- Affordances shape possibilities without determining behaviour.

Kidlink as a Networked Public

FROM A SHORT EXPERIMENT TO A GLOBAL DIALOGUE

- Kidlink began as a fourteen-day online project connecting 260 children in three countries ([Mizell et al., 1992, p. 5](#)).
- The next annual project connected about 2,600 children in 31 countries ([Mizell et al., 1992, p. 6](#)).
- Successive annual projects turned a temporary event into an ongoing global dialogue.
- Separate forums supported conversation, self-description and collective action.
- Internet Relay Chat provided live text conversation in named channels.
- In 1995, children used Kidlink to speak with Zlata Filipović, a Bosnian teenager known for her wartime diary.

From the [raw Kidlink IRC log](#):

[Zlata] To everyone: is Bosnia still on TV? Are people still talking about Sarajevo?

[JodyGee] Zlata: there is a lot of talk about Bosnia lately

[A-Kim] Zlata: Sarajevo was on the front page a few days ago in Toledo, Ohio, USA

([KIDLINK, 1995](#))

Kidlink Changed Its Affordances

DURATION, ACCESS, DISTRIBUTION, AND SEARCH

- **Persistence:** stored messages and annual projects outlasted individual exchanges.
- **Visibility:** child and adult forums created different audiences.
- **Spreadability:** mailing lists, newsletters and printouts moved contributions into other settings.
- **Searchability:** children's introductions became a searchable database.
- Kidlink's organisers closed a peace forum while the technology still worked.
- Changing an affordance changed what the public could become.

Why This Matters

FROM TECHNICAL FEATURE TO SOCIAL CONSEQUENCE

- Persistence helped Kidlink outlast its first experiment.
- Access rules shaped who could speak, observe and organise.
- Distribution and search brought children's contributions to later audiences.
- A feature matters when it changes participation and its consequences.

One Act Becomes Many Things

RELATIONSHIP, RECORD, ARCHIVE, PROFILE, CORPUS

Every single person will be closely followed all his life by his own data shadow where everything he has ever done, learnt, bought, achieved, or failed to achieve is expressed in binary numbers for eternity. His shadow will often be taken for himself and we will have to pay for all the inaccuracies as if they were his own; he will never know who is looking at [his] shadow or why.

Kerstin Anér, quoted in ([Bellovin, 2021](#))

- A social act can remain meaningful participation while also becoming usable data.
- A platform can use recorded activity to define audiences or license data access.
- An archive can preserve the record after the original event.
- Researchers, advertisers and model developers can use traces differently.
- A data shadow preserves some features and leaves others out.

Persistence Without Searchability

URL-BASED ACCESS AND MODEL-ASSISTED DISCOVERY

- Archive.org kept Kidlink pages long after the live site changed.
- The pages survived, but the archive did not make their text searchable.
- Nearly 300,000 Kidlink text records are too many to browse.
- Later research made different parts of the archive visible.
- A model created another partial way into the archive.
- This was AI-assisted classification, not chatting or model training.

The Same Record Serves Different Actors

DISTRIBUTING BENEFIT, COST, RISK, AND CONTROL

- Participants used Kidlink to build international relationships and act together.
- Kidlink stored contributions so the public could continue across later projects.
- Archivists and researchers gained records that preserve history and support new knowledge.
- The people recorded could not control every later audience or use.
- Opportunities and risks differ by stakeholder.

Social Steganography

VISIBLE CONTENT AND UNEQUAL ACCESS TO MEANING

hiding messages in plain sight by leveraging shared knowledge and cues embedded in particular social contexts

([boyd, 2014, p. 65](#))

- Social steganography uses shared context to hide meaning in content visible to several audiences.
- “The Iranian yogurt is not the issue here” became Reddit shorthand for fixating on a detail while missing the larger issue.
- Reddit can record, search and monetise activity without sharing the community’s interpretation.

Advertising Changed Online

FROM PAGE CONTEXT TO INFERRED AUDIENCES

- Contextual advertising matches an advertisement to page content. Audience targeting matches it to people grouped by earlier activity.
- **Persistence:** Reddit can use recent subscriptions and visits to define an audience.
- **Visibility:** public posts, comments, usernames and profiles are widely accessible.
- **Spreadability:** reposting moves eligible content between communities.
- **Searchability:** users find content, while advertisers work from classified audiences.

I Didn't Set a Cookie

A PARTIAL PROFILE FROM BROWSER-EXPOSED SIGNALS

- Rejecting cookies blocks one way of recognising a browser.
- Some signals can be arranged into the form of an advertising bid request.
- Platforms, data providers and advertisers can use partial profiles to define an audience and price access to it.

Displayed demonstration: **Kuber Studio's "I didn't set a cookie" page**

(**Mehta, 2026**)

Why This Matters

FROM ADDRESSABLE AUDIENCE TO LICENSABLE CORPUS

- Reddit participation and browser signals can become usable data.
- Advertisers and platforms act through partial representations of audiences.
- A partial profile omits participants' meanings but still shapes what happens to them.
- Platforms can commercialise access to audiences and to accumulated conversation.

Some Participation Becomes Training Material

REDDIT, DATA LICENSING, AND “AUTHENTIC” CONVERSATION

- Communities created Reddit’s public conversations through years of participation.
- Public pages and Reddit’s earlier data-access service made much of that conversation collectable.
- Reddit now controls structured access and charges for broader commercial use.
- “Authentic human conversation” is a corporate sales claim.

Corpora Are Constructed

COMMON CRAWL AND LICENSED PLATFORM ACCESS

- A training corpus is assembled from selected sources.
- Common Crawl stores dated captures from part of the public web.
- A licensed platform API can provide structured and current access.
- Every collection route includes some material and omits other material.

Training Changes Model Parameters

STATISTICAL PATTERNS ACROSS MANY EXAMPLES

- One selected contribution enters a corpus alongside many others.
- Training repeatedly asks a model to predict the next piece of text.
- Prediction errors adjust numerical settings called parameters.
- Repeated updates encode aggregate statistical patterns.

Training Is Not Chatting

UPDATING PARAMETERS VERSUS GENERATING RESPONSES

- Training changes parameters. Chatting generates responses with existing parameters.
- A conversation supplies temporary context.
- A provider can store chats or memories without changing the base model during the conversation.
- Telling a chatbot something does not ordinarily teach the base model for future users.

Why This Matters

FROM SUPPORTED CLAIM TO DEFENSIBLE RESPONSE

- Reddit now controls current structured access, while earlier external datasets remain outside that route.
- Evidence supports only the process or change that a source documents.
- A defensible response names an actor with control and the consequence their action should change.

When Models Train on Generated Data

COLLAPSE, CURATION, AND THE MYTH OF PURITY

- Generated material can return to online platforms and later training corpora.
- Repeated replacement with generated material can amplify errors and remove uncommon patterns.
- Synthetic data can help when they are selected, mixed and tested.
- The evidence does not support a simple human-good, synthetic-bad rule.

Is the Internet Dead?

BOT TRAFFIC, GENERATED CONTENT, AND HUMAN PRESENCE

- Dead Internet Theory claims that apparent online social life is increasingly synthetic.
- A platform can present people, bots and generated propaganda through the same interface.
- Traffic estimates do not count people or publics.
- Evidence of automation does not establish that a public is synthetic.

What Changes the Relationship?

SYMBOLIC, TECHNICAL, INSTITUTIONAL, AND MARKET RESPONSES

- A “NO AI TRAINING” notice expresses refusal and signals affiliation.
- A community norm does not itself change access, rules or authority.
- Communities can alter collected data, restrict access, seek legal force, leave or fund alternatives.
- A response can be judged by what it changes, for whom and at whose cost.

A contemporary “NO AI TRAINING” notice

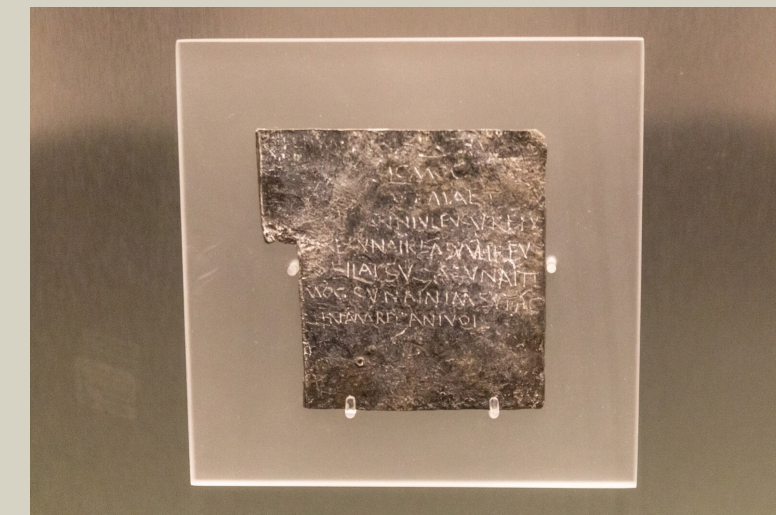
NO AI TRAINING: Without in any way limiting the author's and publisher's exclusive rights under copyright, any use of this publication to “train” generative artificial intelligence (AI) technologies to generate text is expressly prohibited. The author reserves all rights to license uses of this work for generative AI training and development of machine learning language models.

It is a very odd feeling seeing history repeat itself
cause back in *my* day, guide writers on gamefaqs would plaster copyright notices at the top of their faqs, in a vain attempt to get scrapers to stop stealing their work
and as the months/years went by their 'copyright notices' would get increasingly lengthy and unhinged, yelling into the void
because it turns out that whether it's the 90's or today
content scrapers give zero fucks about legality

A contemporary notice prohibiting use of a publication for AI training

Private Discord correspondence, reproduced with permission (personal communication, 13 August 2026)

A Roman curse tablet



A metal Roman curse tablet complaining about the theft of Vilbia

May he who has stolen VILBIA from me become as liquid as water ...

([Roman Inscriptions of Britain, n.d.](#))

Tablet photograph by -JvL-, CC BY 2.0 ([-JvL-, 2022](#))

Books Become Data

TEXT PRESERVED, PHYSICAL COPIES DESTROYED

- In *Rainbows End*, a library preserves the text by destroying the books.
- 404 Media used an AirTag to follow a book to an Amazon facility that scans books for AI training.
- The report does not show what happened to the tagged book.

[404 Media, “We Tracked a Shipment of Rare Books,” 17 August 2026](#)

What Does “Rare” Mean?

A HEADLINE WITHOUT THE TITLES

- 404 Media calls the tracked shipment “rare books” but does not name them.
- Earlier bulk orders included *The Insider’s Guide to Metro Denver* (1995) and *How to Use Corel WordPerfect* (1991).
- Becker called books like these dead inventory and “not necessarily rare.”
- Rarity depends on importance and demand, not scarcity alone.
- The headline can circulate while the books needed to assess it remain unidentified.

[404 Media, 17 August 2026](#) · [Charlie Becker, 10 May 2026](#) · [RBMS, “Your Old Books”](#)

What If Someone Brings a Very Big Cheque?

OUR ASSESSMENTS ARE DATA

- University systems accumulate assignments, drafts, feedback and forum posts.
- These records show how people learn, not only what they submit.
- Weller imagines companies selling access to this corpus for AI training.
- What happens when someone arrives with a very big cheque for all of it?

[Martin Weller, “Slop Will Eat Itself \(then us\),” 17 August 2026](#)

References

umgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). *The Pushshift Reddit Dataset*. <https://doi.org/10.48550/ARXIV.2001.08435>

llovin, S. M. (2021, June 26). *Where Did “Data Shadow” Come From?* SMBlog. <https://www.cs.columbia.edu/~smb/blog/2021-06/2021-06-26.html>

rch, K., & Cochrane, D. T. (2022). Big Tech: Four Emerging Forms of Digital Rentiership. *Science as Culture*, 31(1), 44–58. <https://doi.org/10.1080/09505431.2021.1932794>

yd, danah. (2014). *It's Complicated: The Social Lives of Networked Teens*. Yale University Press.

icher, T., & Helmond, A. (2018). The Affordances of Social Media Platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE Handbook of Social Media* (pp. 233–253). SAGE Publications Ltd. <https://doi.org/10.4135/9781473984066.n14>

ommon Crawl Foundation. (n.d.). *Overview*. Common Crawl. Retrieved August 15, 2026, from <https://commoncrawl.org/overview>

Haas, M. (1995, June 2). *Zlata Chat (Edited Version)*. KIDLINK. <https://web.archive.org/web/20010512082947/http://www.kidlink.org:80/KIDPROJ/WritersCorner/EVENTS/zlata.html>

ites, K. (2025). “A High-Tech Company Masquerading as a Retailer”: Target's Video Infrastructure. In *Targeted: Corporations and the Police Surveillance Economy* (pp. 52–84). New York University Press. <https://www.jstor.org/stable/jj.27300020.6>

rstgrasser, M., Schaeffer, R., Dey, A., Rafailov, R., Sleight, H., Hughes, J., Korbak, T., Agrawal, R., Pai, D., Gromov, A., Roberts, D. A., Yang, D., Donoho, D. L., & Koyejo, S. (2024). *Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data*. <https://doi.org/10.48550/ARXIV.2404.01413>

oogle. (n.d.). *Contextual Targeting*. Google Ads Help. Retrieved August 16, 2026, from <https://support.google.com/google-ads/answer/1726458>

uminatiPirate. (2021, January 5). *Dead Internet Theory: Most of the Internet Is Fake* [Forum post]. Agora Road's Macintosh Cafe. <https://forum.agoraroad.com/index.php?threads/dead-internet-theory-most-of-the-internet-is-fake.3011/>

hnston, V. H., Ballsun-Stanton, B., Jensen, H. S., Kjelsen, C. K., & Thøgersen, J. (2026). *Exploring the Archived Web through AI Assisted Document Discovery* (Preprint bc77ead95211793cbaaebdfce200e50e6bbb365d). <https://github.com/WEB-CHILD/exploring-the-archived-web-through-ai-assisted-document-discovery> (Pre-published)

vL-. (2022). *A Metal Curse Tablet (Defixio) with a Complaint about the Theft of a Vilbia* [Graphic]. [https://commons.wikimedia.org/wiki/File:A_metal_curse_tablet_\(defixio\)_with_a_complaint_about_the_theft_of_a_Vilbia.jpg](https://commons.wikimedia.org/wiki/File:A_metal_curse_tablet_(defixio)_with_a_complaint_about_the_theft_of_a_Vilbia.jpg)

Kagi. (n.d.). *Pricing*. Kagi Search. Retrieved August 16, 2026, from <https://kagi.com/pricing>

Karpathy, A. (Director). (2023, November 23). *[1hr Talk] Intro to Large Language Models* [Video recording]. https://www.youtube.com/watch?v=zjKBMFhNj_g

KIDLINK. (1995, June 2). *Zlata Filipovic on KIDLINK IRC*. <https://web.archive.org/web/20001117135000/http://www.kidlink.org:80/resources/irczlata.html>

Klincewicz, M., Alfano, M., & Fard, A. E. (2025). Slopaganda: The interaction between propaganda and generative AI. *Filosofiska Notiser*, 12(1), 135–162. <https://www.filosofiskanotiser.com/KlincewiczAlfanoFard.pdf>

Laperdrix, P., Bielova, N., Baudry, B., & Avoine, G. (2020). Browser Fingerprinting: A Survey. *ACM Transactions on the Web*, 14(2), 8:1–8:33. <https://doi.org/10.1145/3386040>

Mehta, K. (2026, August 5). *I Didn't Set a Cookie*. Kuber Studio. <https://kuber.studio/cookie/>

Mizell, A. P., Carroll, P., & Haddad, V. (1992). *Global Communications: Kids-92* (Reports – Descriptive; Speeches/Meeting Papers No. ED368342; p. 19). <https://eric.ed.gov/?id=ED368342>

Ng, L. H. X., & Carley, K. M. (2025). A global comparison of social media bot and human characteristics. *Scientific Reports*, 15(1), 10973. <https://doi.org/10.1038/s41598-025-96372-1>

OpenRTB 2.6, Technical specification Nos. OpenRTB 2.6 (2022). <https://github.com/InteractiveAdvertisingBureau/openrtb2.x/blob/main/2.6.md>

Patel, R. (2024, February 22). *An Expanded Partnership with Reddit*. Google. <https://blog.google/company-news/inside-google/company-announcements/expanded-reddit-partnership/>

Prada, L. (2025, September 15). *One-Third of the Internet Is Just Bots Now. Seriously*. VICE. <https://www.vice.com/en/article/yep-one-third-of-the-internet-is-just-bots-now/>

rAmltheAsshole. (2019, May 1). *AITA for Throwing Away my Boyfriend's Potentially Illegal Yogurt Collection?* [Reddit post]. https://www.reddit.com/r/AmltheAsshole/comments/bjd41e/aita_for_throwing_away_my_boyfriends_potentially/

rAmltheAsshole. (2022a, October 20). *The Iranian yogurt is not the issue here*. [Reddit comment]. https://www.reddit.com/r/AmltheAsshole/comments/y8gzd5/aita_for_sitting_on_my_wifes_bag_after_she/it0t39y/

rAmltheAsshole. (2022b, October 20). *What's the reference for the Iranian yogurt, I haven't seen the post. Can someone please link it* [Reddit comment]. https://www.reddit.com/r/AmltheAsshole/comments/y8gzd5/aita_for_sitting_on_my_wifes_bag_after_she/it1ghz3/

Rands. (2026, August 13). *RIP Claude*. Rands in Repose. <https://randsinrepose.com/archives/rip-claude>

Reddit. (n.d.). *Community and Interest Targeting*. Reddit for Business. Retrieved August 13, 2026, from <https://www.business.reddit.com/advertise/targeting/community-and-interest>

Reddit. (2024, February 22). *Expanding Our Partnership with Google*. Reddit. <https://redditinc.com/news/reddit-and-google-expand-partnership>

Reddit. (2025, May 29). *Public Content Policy*. Reddit Help. <https://support.reddithelp.com/hc/en-us/articles/26410290525844-Public-Content-Policy>

Reddit. (2026a, May 26). *How Does Reddit Search Work?* Reddit Help. <https://support.reddithelp.com/hc/en-us/articles/19695647891988-How-does-Reddit-search-work>

Reddit. (2026b, May 28). *Developer Platform and Accessing Reddit Data*. Reddit Help. <https://support.reddithelp.com/hc/en-us/articles/14945211791892-Developer-Platform-Accessing-Reddit-Data>

Reddit. (2026c, July 13). *What Is Reposting (FKA Crossposting)?* Reddit Help. <https://support.reddithelp.com/hc/en-us/articles/4835584113684-What-is-reposting-fka-crossposting>

Roman Inscriptions of Britain. (n.d.). *Tab. Sulis 4: Curse Tablet* [Inscription]. Roman Inscriptions of Britain. Retrieved August 16, 2026, from <https://romaninscriptionsofbritain.org/inscriptions/TabSulis4>

Ross, S., & Ballsun-Stanton, B. (2026, July 24). *Paper B: Reliability in research with large language models is a property of the human–AI system*. <https://doi.org/10.17605/OSF.IO/M376W>

Shan, S., Ding, W., Passananti, J., Wu, S., Zheng, H., & Zhao, B. Y. (2023). *Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models*. <https://doi.org/10.48550/ARXIV.2310.13828>

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755–759. <https://doi.org/10.1038/s41586-024-07566-y>

The Prisoner (1967): *HD Opening Titles*. (2024, March 15). [Video recording]. https://www.youtube.com/watch?v=0a1_v-BJ-Wc

Willison, S. (2024, May 29). *Training is not the same as chatting: ChatGPT and other LLMs don't remember everything you say*. Simon Willison's Weblog. <https://simonwillison.net/2024/May/29/training-not-chatting/>



AI-use disclosure

- Claude (Anthropic) and Codex (OpenAI) produced about 4.3 million output tokens between 11 and 18 August.
- About 2.7 million tokens came from main threads and 1.6 million from subagents.
- The tools supported source research, planning, drafting, production and verification.
- Brian selected the argument, sources and teaching form.