

Digital Literacy & Crowdsourcing: Tackling Descriptive and Algorithmic Bias through Doing

Dr. Jessica BrodeFrank, Clinical Assistant Professor & Digital Librarian
University of Illinois Chicago

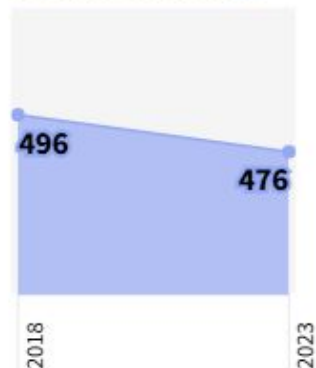
What is digital literacy?

the capabilities which fit
someone for living,
learning, and working in a
digital society - JISC

Digital Literacy in Crisis?

Computer and information literacy and computational thinking scores

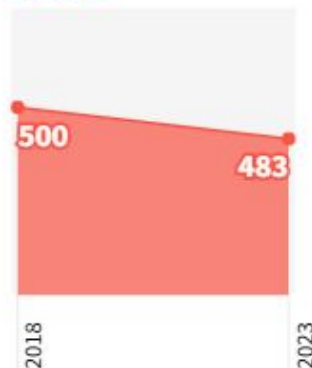
Computer and information literacy score: international average



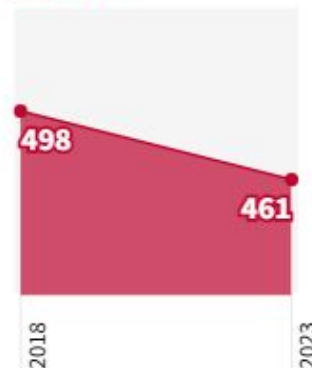
Computer and information literacy score: US



Computational thinking score: international average



Computational thinking score: US



DATA SOURCE: NCES



*A Flourish chart

How did we get here?

- Digital divide
- Lack of digital and information literacy training
- Generative AI



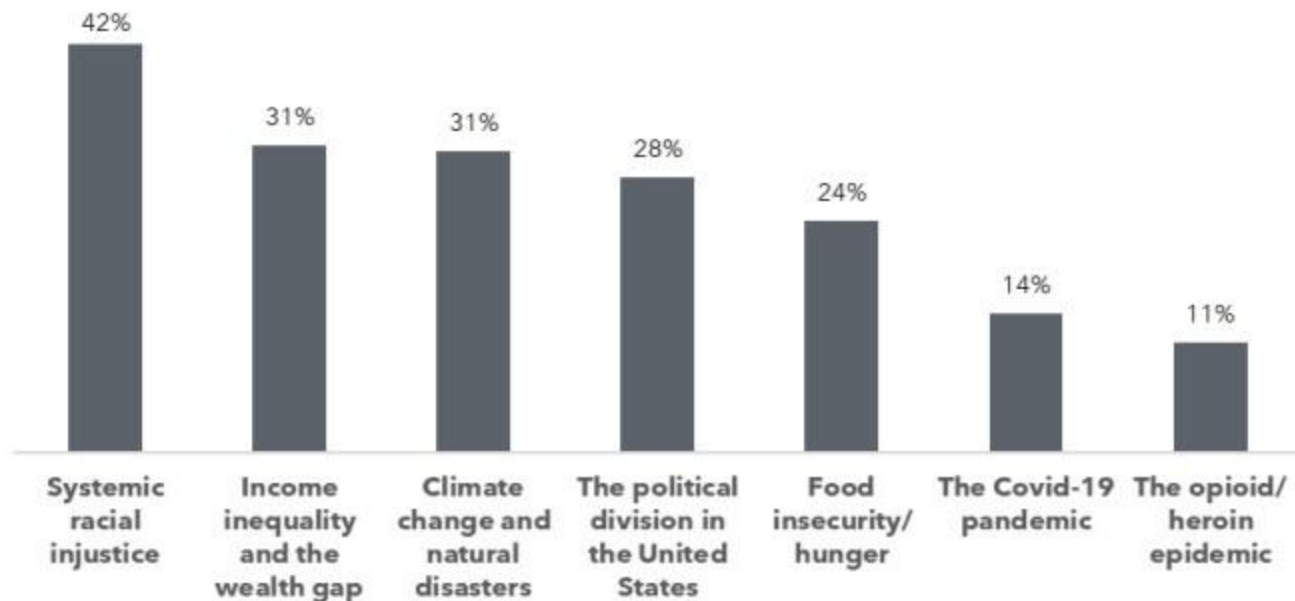
Information Literacy Resources for Library and Museum Professionals

Information literacy is a set of skills that people need to find, understand, evaluate, and share information. You can use the resources on this site to plan information literacy programs and help community members build these critical skills.

Start your community's journey to information literacy



Fig. 9 | Social issues Americans want culture organizations to address. "Which of the following social issues (if any) do you think arts or culture organizations should address? Please select up to THREE or tell us in your own words." Wave 2 only, n = 51,702. (q48)



<https://search.issuelab.org/resource/rethinking-relevance-rebuilding-engagement.html>

Libraries and museums as sites of trusted
information sources, public engagement
experience, and education expertise for sites
for digital literacy

The American Alliance of Museums (AAM) Code of Ethics for Museums stated that:

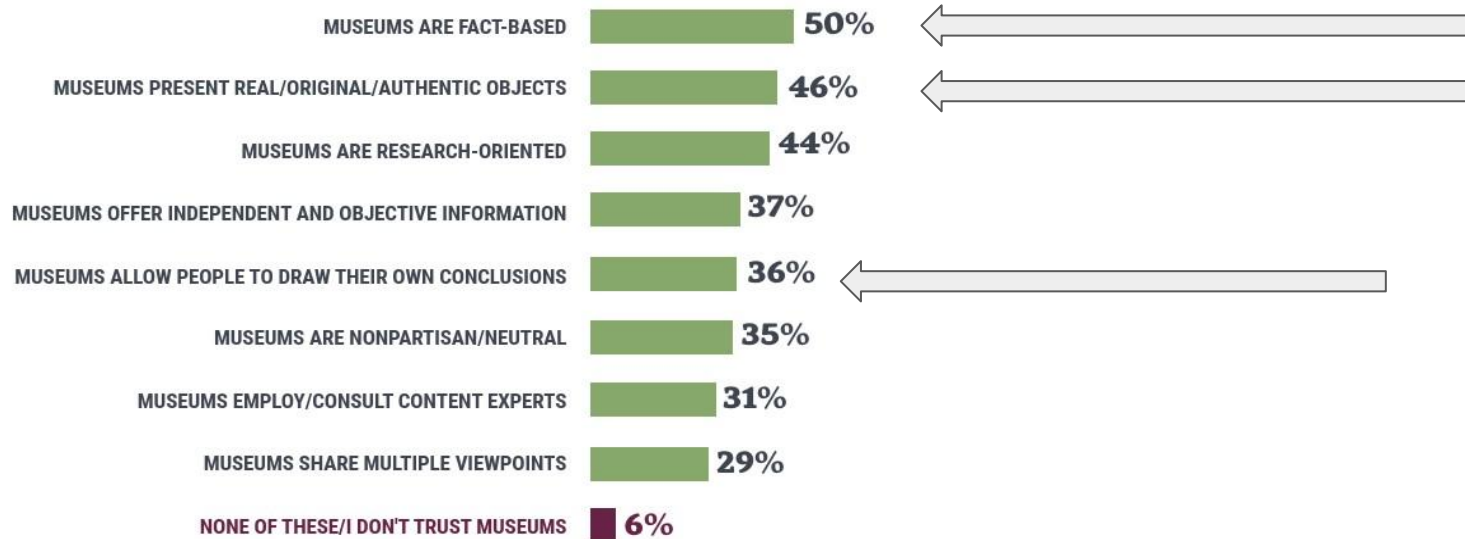
“Museums in the United States are grounded in the tradition of public service. They are organized as public trusts, holding their collections and information as a benefit for those they were established to serve.”

American Alliance of Museums. “Public Trust and Accountability Standards,” January 4, 2018.

<https://www.aam-us.org/programs/ethics-standards-and-professional-practices/public-trust-and-accountability-standards/>.



Reasons people gave that **contribute** to trust in museums:



American Alliance of Museums. “Museum and Trust 2021,” September 30, 2021.
<https://www.aam-us.org/2021/09/30/museums-and-trust-2021/#:~:text=Museums%20consistently%20rank%20among%20the,a%20similar%20level%20of%20trust>.

Introduction to Crowdsourcing

At its best, [crowdsourcing] is a form of digitally-enabled participation that creates meaningful opportunities for the public to experience collections while undertaking a range of tasks that make those collections more easily discoverable and usable by others.

- Ridge, Blickhan, Ferriter et al. (2022) *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage*

Why crowdsource?

Discoverability	Engagement
Data processing/transformation	Opportunities for exploration
Increased access to collections	Encourage institutional transparency around the digitization process
Results in new, or expanded, reusable datasets	Work with communities toward a shared goal

Importance of Metadata

Tammy Lee, Culture Creates in Montreal Canada, stated in March 2025, metadata needs to be seen as more than just data about data, **“it’s the connective tissue of the internet – the invisible layer that decides what gets seen, what gets connected, and what gets left out.”**

Metadata makes content legible to machines, it makes content discoverable. Without it, algorithms have nothing to work with. Tagging and metadata creation are the key to amplifying diversity. You cannot surface underrepresented voices if the system doesn’t know they’re there.

Crowdsourcing in libraries & museums

Tag Along with Adler

- Zooniverse Hosted
- 1090 Adler Planetarium Collections Objects
 - 613 Works on Paper
 - 195 Archival Rare Photographs
 - 282 Rare Book Illustrations
- 2 Workflows
 - Tag Images
 - Verify AI Tags
 - AI used: the iMet Collection Attribute Classifier and the Google Cloud Vision API taggers
- Launched publicly in March 2021
 - 11 incremental data releases (10 of 100 images, 1 of 90 images)
- Completed in March 2022

Bridging the Semantic Gap - Tag Along with Adler

322,993

Tags Generated by Volunteers

6,976

Individual Project Participants

Average of
363.4 Tags
Added per
Image

>80% of Tags
Added Were Not
in Adler
Catalogue or AI
Tags

Bridging the Semantic Gap between the Cataloguer and the Public



woman
meteorite
Chicago Park District
1945
1950
Adler Planetarium
Chicago



Asteroid
Blonde
Display
exhibit
Exhibition
Fossil
Girl
meteor
Meteorite
Moon rock

Museum
Planets
Photograph
portrait
pose
Rock
sitting
sitting on
Space rock
vintage
Woman
Young Woman

Table 3. Nature of errors on DDC exercises.

	ChatGPT (n = 17)	Copilot (n = 21)	Gemini (n = 19)
Incorrect topic	5	6	8
DDC number does not exist	2	2	0
DDC number too general	7	12	8
DDC number too specific	1	0	1
Did not follow order rules	2	1	2

Table 4. AI performance on LCC exercises.

	ChatGPT	Copilot	Gemini
Correct	2	5	1
Close	2	0	0
Acceptable alternative	1	1	0
Incorrect	20	19	22
Refusal	0	0	2
Final grade	20%	24%	4%
Number of valid LCCs	13	13	3
Percentage valid	52%	52%	13%

Table 5. Nature of errors on LCC exercises.

	ChatGPT (n = 20)	Copilot (n = 19)	Gemini (n = 22)
Incorrect topic	6	12	11
LCC number does not exist	2	0	8
LCC number too general	6	5	1
LCC number too specific	2	0	0
Provided main class only	3	2	2
Provided number range only	1	0	0

Table 2. AI performance on DDC exercises.

	ChatGPT	Copilot	Gemini
Correct	5	1	4
Acceptable alternative	1	1	0
Incorrect	17	21	19
Refusal	0	0	0
Final grade	26%	9%	17%
No. of valid DDCs	16	14	16
Percentage valid	70%	61%	70%

Dobreski, Brian, and Christopher Hastings. “AI Chatbots and Subject Cataloging: A Performance Test.” Library Resources & Technical Services 69, no. 2 (April 9, 2025). <https://doi.org/10.5860/lrts.69n2.8440>.

Generative AI ≠ Ontological Search

28% of Americans state they use generative AI tools at least weekly.

14% use generative AI tools for search **daily**.

Ballard, Jamie. “Americans Are Increasingly Skeptical about AI’s Effects | YouGov,” March 13, 2025.

<https://today.yougov.com/technology/articles/51803-americans-increasingly-skeptical-about-ai-artificial-intelligence-effects-poll>.

AI models incorrectly cited sources in more than **60 percent** of their queries

A.I.-Generated Reading List in Chicago Sun-Times Recommends Nonexistent Books

A summer reading insert recommended made-up titles by real authors such as Isabel Allende and Delia Owens. The Sun-Times and The Philadelphia Inquirer have apologized.

Edwards, Benj. “AI Search Engines Cite Incorrect News Sources at an Alarming 60% Rate, Study Says - Ars Technica,” March 13, 2025.

Unique collections and opportunities to tackle current threats to digital literacy

AI Algorithmic Bias

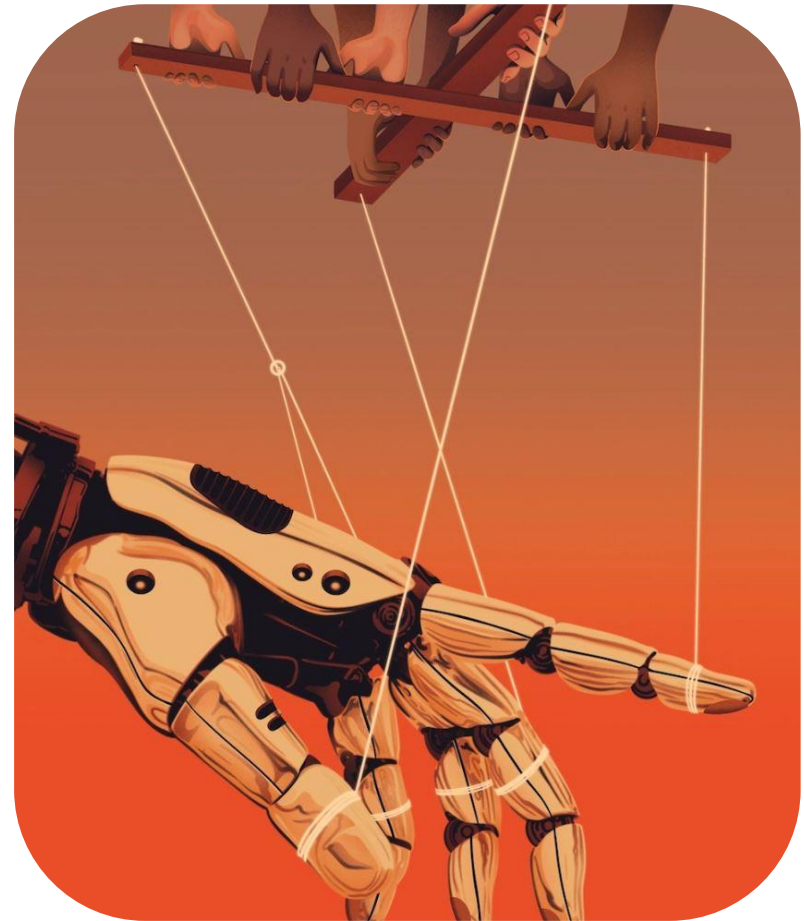
"how dangerous it is for machine learning systems to be in the business of 'classifying' humans and how easily efforts can - and do - go horribly wrong." artist Trevor Paglen and AI researcher, Kate Crawford

"Humans categorized what they saw in terms of race, gender, age, emotion, and sometimes personal character. In so doing, they injected their own conscious and unconscious opinions and biases into the tissue of the algorithm."

artnet News. "How ImageNet Roulette, a Viral Art Project That Exposed Facial Recognition's Biases, Is Changing Minds About AI," September 23, 2019. <https://news.artnet.com/art-world/imagenet-roulette-trevor-paglen-kate-crawford-1658305>.

Ethics in AI - Labor Ethics

OpenAI used outsourced Kenyan laborers earning less than \$2 per hour, a TIME investigation has found. – January 18, 2023.

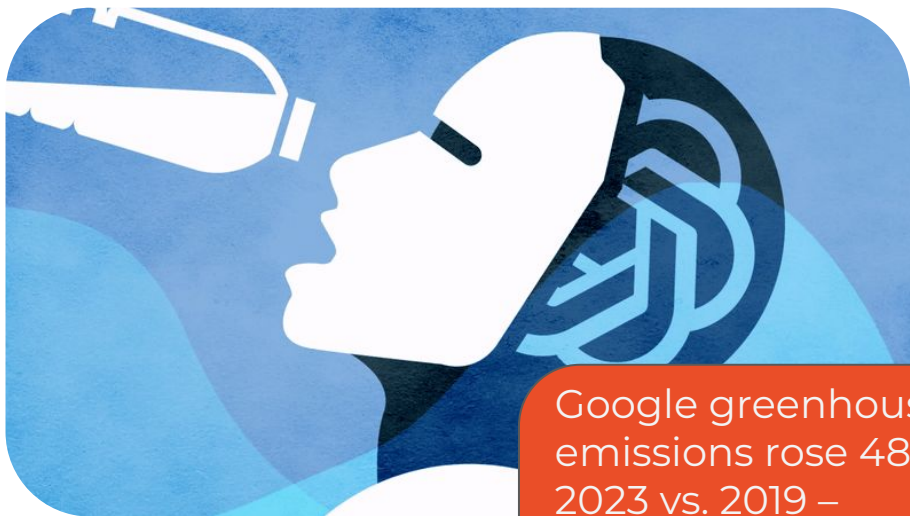


Nash Weerasekera for Noema Magazine:
<https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/>

Ethics in AI - Climate Concerns

4.32 grams of CO2 emissions -
average Chat GPT query

1.69g of CO2 emissions -
Zooniverse crowdsourcing platform hit



<https://www.thetimes.com/uk/technology-uk/article/hirsty-chatgpt-uses-four-times-more-water-than-previously-thought-bc0pqswd>

Google greenhouse gas emissions rose 48% in 2023 vs. 2019 – attributed to data center energy consumption. – 7/12/2024 NPR

