

Some things to consider when
deciding whether to start
building with “AI” in libraries.

AI?

- I'm a software developer who dabbles in data science, not an artificial intelligence or machine learning expert.
- I tend to prefer the term **Machine Learning** to Artificial Intelligence (AI) since it speaks to the trajectory of computing in the last decade.
- I'll be talking about recent development of Large Language Models (LLM) like ChatGPT, but some things apply more broadly.
- AI Boomers and Doomers are two sides of the same coin. It's important to think critically and to remember "It could be otherwise".
- McQuillan, D. (2022). [Resisting AI: An Anti-facist Approach to Artificial Intelligence](#).



Bias

LLMs are built large amounts content collected from the web.

The web is a big place so decisions have been made about what is used.

Corporations are increasingly tight lipped about where there training data comes from.

LLM content is being made part of the web and leads to “model collapse”.

Tactic:

When evaluating an AI tool see if you can determine what content was used to train the model.



Intellectual Property

LLMs have been trained with copyrighted content from the web.

The Authors Guild and the New York Times are suing OpenAI.

Chat style generative AI breaks the web ecosystem because they don't link to websites.

Organizations are starting to block Google AI, OpenAI and CommonCrawl.

Tactic:

What content should we make available to Generative AI tools. What would our donors want?



Verifiability

LLMs don't record the provenance of their data, and aren't able to cite sources.

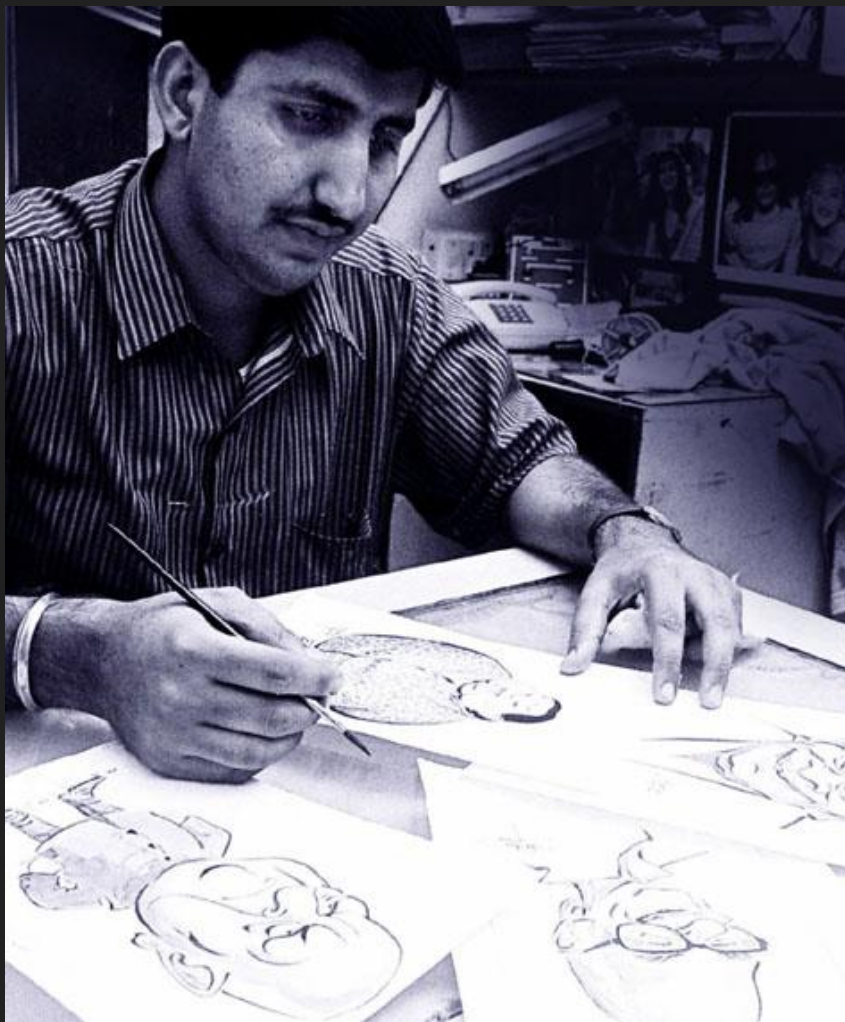
Model explainability is an active research area and some aren't sure it's possible.

Research has found that generative AI tools that appear to cite are unreliable.

Linking to source documents is foundational to what libraries do.

Tactic:

Library and Archives professionals have a role in evaluating how AI tools cite documents as evidence.



Work

AI tools claim to democratize access to skills: programming, illustration, writing, etc.

AI tools threaten to replace or more likely deskill professions.

AI tools use “ghost workers” and RLHF to outsource and underpay.

Tactic:

When evaluating the use of AI tools involve the people whose work is impacted in the decision making and implementation.



Sustainability

It cost \$100 million to train GPT-4.

GPUs use 4 times as much energy as CPUs.

Water used for cooling data centers is also an issue.

Querying LLMs takes is estimated to use 60-100 times more energy than traditional search.

Jevons Paradox: efficiency gains are lost to increased demand.

Tactic:

We should be looking for ways to reduce energy consumption, not increase it.



Security & Privacy

Generative AI is being used by bad actors to “flood the zone” with disinformation.

Lack of explainability means that LLMs are vulnerable to “prompt injection”.

Tools that integrate user data with AI via APIs leak data that could be used to further train models.

Tactic:

Support tools and standards that provide transparency, authenticity and give users agency over their data.

Sources: Bias

Baack, S. (2024). [Training Data for the Price of a Sandwich: Common Crawl's Impact on Generative AI](#). Mozilla Foundation.

Dupré, M. (2024). [Wikipedia No Longer Considers CNET a "Generally Reliable" Source After AI Scandal](#). Futurism.

Shumailov, I et al (2023). [The Curse of Recursion: Training on Generated Data Makes Models Forget](#). arXiv.

Eubanks, V. (2018). [Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor](#). Macmillan.

Sources: Intellectual Property

Reisner, A. (2023). [Revealed: The Authors Whose Pirated Books are powering Generative AI.](#) The Atlantic.

David, Emilia. (2024). [OpenAI's news publisher deals reportedly top out at \\$5 million a year.](#) The Verge.

Grynbaum, M. and Mac, R. (2023). [The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work.](#) The New York Times.

Welsh, B. (2023). [Who blocks OpenAI, Google AI and Common Crawl.](#) Palewire.

Sources: Verifiability

Zhao, H. et al. (2023). [Explainability for Large Language Models: A Survey](#). ACM Transactions on Intelligent Systems and Technology.

Liu, N. et al. (2023). [Evaluating Verifiability in Generative Search Engines](#). arXiv.

Zuckerman, E. (2023). [Heather Ford: Is the Web Eating Itself? LLMs versus verifiability](#).

Sources: Work

Crabapple, M. (2022). [Beware a world where artists are replaced by robots.](#) It's starting now. Los Angeles Times.

[Coding on Copilot: 2023 Data Suggests Downward Pressure on Code Quality.](#) GitClear.

Xiang, C. (2023). [GPT-4 Can't Replace Striking TV Writers, But Studios Are Going to Try.](#) Vice.

Mateescu, A. (2024). [New Technologies Require Thinking More Expansively About Protecting Workers.](#) Tech Policy Press.

Roberts, S. (2021). [Behind the Screen: Content Moderation in the Shadows of Social Media.](#) Yale University Press.

Sources: Sustainability

Payton, B. (2023). [Power mad: AI's massive energy demand risks causing major environmental headaches.](#) Reuters.

Crawford, K. (2024). [Generative AI's environmental costs are soaring — and mostly secret.](#) Nature.

Vanderbauwhede, W. (2023). [The climate cost of the AI revolution.](#)

Verdecchia, R. (2023). [A systematic review of Green AI.](#) WIREs.

De Vries, A. (2023). [The growing energy footprint of artificial intelligence.](#) Joule.

2024. [Markey, Heinrich, Eshoo, Beyer Introduce Legislation to Investigate, Measure Environmental Impacts of Artificial Intelligence.](#)

Sources: Security & Privacy

Cohen, S. et al (2024). [Here Comes The AI Worm: Unleashing Zero-click Worms that Target GenAI-Powered Applications.](#) arXiv.

Willison, S. (2023). [Prompt injection: What's the worst that can happen?](#)

Barrón-López, L. (2024). [How AI-generated misinformation threatens election integrity.](#) PBS News Hour.

Weisner, B. (2023). [Here's What Happens When Your Lawyer Uses ChatGPT.](#) New York Times.

Kan, M. [DocuSign Tapping User Data to Train AI Models, Offers Vague Privacy Promises.](#) PC Magazine.

Sources: Images

Every brain is unique. https://commons.wikimedia.org/wiki/File:Neurodiversity_Crowd_1.png

Walls and fields

https://commons.wikimedia.org/wiki/File:Walls_and_fields_-_geograph.org.uk_-_1857971.jpg

Gold Twitter verified badge

https://commons.wikimedia.org/wiki/File:Twitter_Verified_Badge_Gold.svg

A cartoonist / Illustrator at work https://commons.wikimedia.org/wiki/File:Cartoonist_at_work.jpg

Aerial view of the Google Data Center in Council Bluffs, IA.

[https://commons.wikimedia.org/wiki/File:Google_Data_Center,_Council_Bluffs_Iowa_\(49062863796\).jpg](https://commons.wikimedia.org/wiki/File:Google_Data_Center,_Council_Bluffs_Iowa_(49062863796).jpg)

'Desinformación', mural de Lalo Luque (Lalone) de octubre del año 2020 en el número 33 de la calle Antonio Vico, Ciudad Jardín, Málaga, España.

[https://commons.wikimedia.org/wiki/File:Desinformaci%C3%B3n_\(Lalone,_2020\).jpg](https://commons.wikimedia.org/wiki/File:Desinformaci%C3%B3n_(Lalone,_2020).jpg)