

Academic libraries in an Open Access and AI first world – an attempt to peer into the future

Aaron Tay (Lead, Data Services)

Singapore Management University Libraries

43rd IATUL Conference UAE 2023



Aarontay@gmail.com



[@aarontay](https://twitter.com/aarontay)



<https://musingsaboutlibrarianship.blogspot.com>

ChatGPT's attempt to keynote



Dear esteemed colleagues,



I am thrilled to stand before you today and discuss how the advancement of technology is going to radically change academic libraries. Specifically, I want to talk about how the rise of Generative AI, powered by technologies like large language models and text-to-image AI, is going to impact academic libraries in a significant way.

As we all know, technology is rapidly advancing, and AI is becoming more sophisticated every day. But what is even more exciting is that academic libraries are going to be impacted even more. Why is that? It's because even the most advanced algorithm will not work without data to train on.

The good news is that we are currently in the perfect storm. On one hand, the scholarly ecosystem is now surfacing, collecting and making available more than just final outputs like journal articles and books. Instead, we are now looking at collecting and discovering data, protocols, code, preprints, peer reviews, and more. We no longer stop when an output is published; we now study post-publication reviews, retractions, and other elements.

On top of that, the rise of movements like Open Science, Open Data, Open Access, Open Metadata, and Open Infrastructure is creating even more data to train on. This presents a tremendous opportunity for academic libraries. We can play a significant role in ensuring that AI algorithms have access to high-quality training data, helping to improve their accuracy and effectiveness.

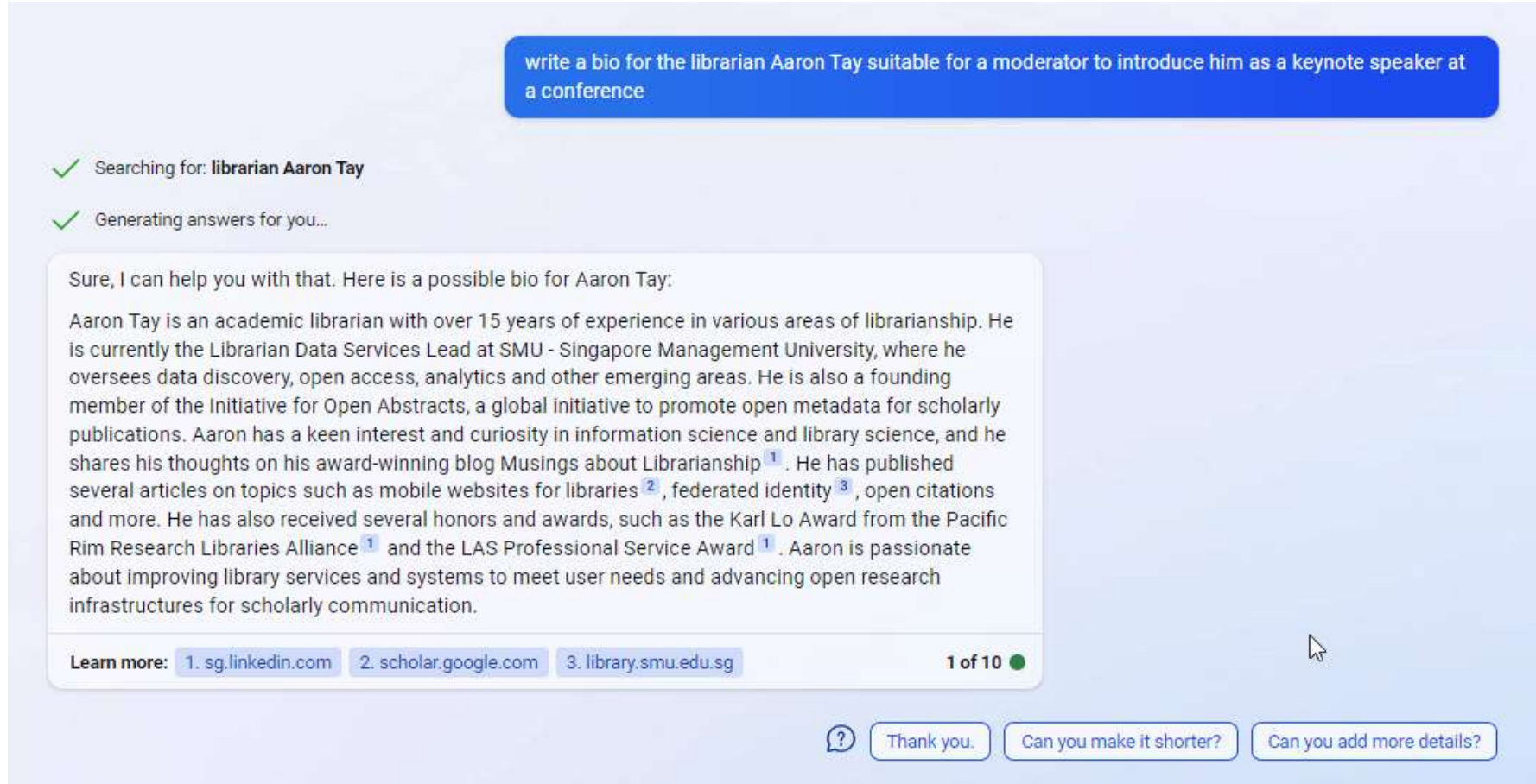
However, this also means that academic libraries need to be proactive in how we manage data. We need to ensure that we are collecting and curating data in a way that is useful for AI algorithms. We need to think about how we can make this data more accessible to researchers, both within and outside our institutions.

We also need to be aware of the ethical considerations around AI and data. There are concerns around bias and fairness, as well as privacy and security. As librarians, we have a responsibility to ensure that the data we provide for AI algorithms is both accurate and ethically sourced.

In conclusion, the rise of Generative AI, powered by technologies like large language models and text-to-image AI, is going to have a significant impact on academic libraries. But this impact is not something to fear; it's something to embrace. We have a unique opportunity to play a significant role in ensuring that AI algorithms have access to high-quality training data, which will help to improve their accuracy and effectiveness. As long as we are proactive in managing data and mindful of ethical considerations, academic libraries will continue to be at the forefront of research and discovery.

Thank you.

Bio written by new Bing (Bing+GPT model)



write a bio for the librarian Aaron Tay suitable for a moderator to introduce him as a keynote speaker at a conference

✓ Searching for: **librarian Aaron Tay**

✓ Generating answers for you...

Sure, I can help you with that. Here is a possible bio for Aaron Tay:

Aaron Tay is an academic librarian with over 15 years of experience in various areas of librarianship. He is currently the Librarian Data Services Lead at SMU - Singapore Management University, where he oversees data discovery, open access, analytics and other emerging areas. He is also a founding member of the Initiative for Open Abstracts, a global initiative to promote open metadata for scholarly publications. Aaron has a keen interest and curiosity in information science and library science, and he shares his thoughts on his award-winning blog Musings about Librarianship¹. He has published several articles on topics such as mobile websites for libraries², federated identity³, open citations and more. He has also received several honors and awards, such as the Karl Lo Award from the Pacific Rim Research Libraries Alliance¹ and the LAS Professional Service Award¹. Aaron is passionate about improving library services and systems to meet user needs and advancing open research infrastructures for scholarly communication.

Learn more: [1. sg.linkedin.com](#) [2. scholar.google.com](#) [3. library.smu.edu.sg](#) 1 of 10 ●

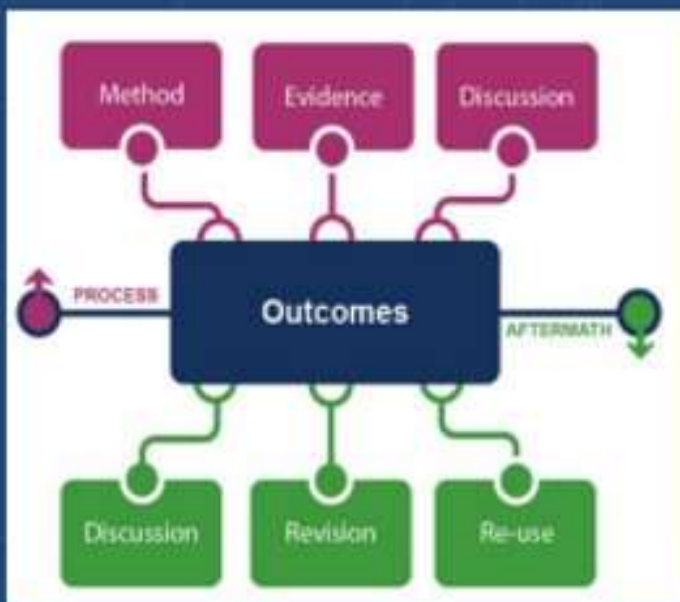
? Thank you. Can you make it shorter? Can you add more details?

Diversity in record + Open + New Technology (e.g. ML) = New roles

Diversity in scholarly record

Push to Open

Technology



Open Access
Open Data
Open Edu resources
Open Citations
Open Science
Open infrastructure



Linked data
Machine Learning
NLP

Open + More stuff + Machine learning = magic

What does the title of the talk mean?

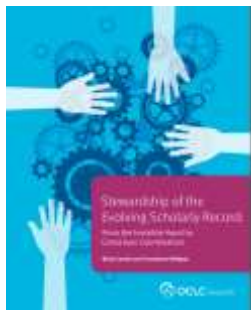
"finally cash the cheques written by the Open Access movement."
-- Our Research, July 2018

We're building the "AI-powered support tools" now. What kind of tools? Well, let's go back to the Hamlet example...today, publishers solve the context problem for readers of Shakespeare by adding notes to the text that define and explain difficult words and phrases. We're gonna do the same thing for 20 million scholarly articles. And that's just the start...we're also working on concept maps, automated plain-language translations (think automatic [Simple Wikipedia](#)), structured abstracts, topic guides, and more. Thanks to recent progress in AI, all this can be automated, so we can do it at scale. That's new. And it's big. -- [OurResearch Blog](#), 2018

Trend 1 – the evolving Scholarly Record (2014)



Figure 1. The Evolving Scholarly Record



<https://www.oclc.org/research/publications/2014/oclcresearch-evolving-scholarly-record-2014-overview.html>

Things we used to collect and care about



BOOKS



JOURNAL
ARTICLES



CONFERENCE
PROCEEDINGS



NEWSPAPERS

Things we used to collect and care about



Preprints (different versions e.g. Accepted Manuscript, Version of Record)



Datasets – raw, processed (text, images and more - See list of [data repositories](#))



Computational Notebooks & Scripts/code
(e.g. [CodeOcean](#), [Executable Research Articles](#))



Protocols (including search protocols), Registered reports (e.g. [SearchRxiv](#), [protocols.io](#), OSF,)



Peer Review reports (open peer review models, [Peer community](#))

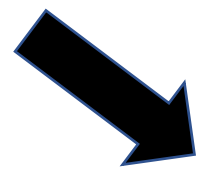


Post & Pre publication reviews (e.g. [Publons](#), [PubPeer](#))



Retraction notices (Crossmark, Retraction Watch Database)

Raw Data



Preprint reviews

Open Peer Review

Post – publication review

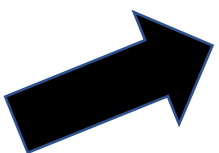
Registered protocol



Preprint	Accepted Manuscript	Version of Record
----------	---------------------	-------------------

Preprint Server	Institutional Repository	Publisher platform
-----------------	--------------------------	--------------------

Code



Discovery, Access and metrics for evaluation

Citations

Usage - Downloads

Usage – Other altmetrics

The preprint revolution - Implications for bibliographic databases

The preprint revolution - Implications for bibliographic databases

How comprehensive are bibliographic databases when it comes to preprints? Ludo Waltman and Nees Jan van Eck provide six recommendations for improvement.

LUDO WALTMAN and NEES JAN VAN ECK

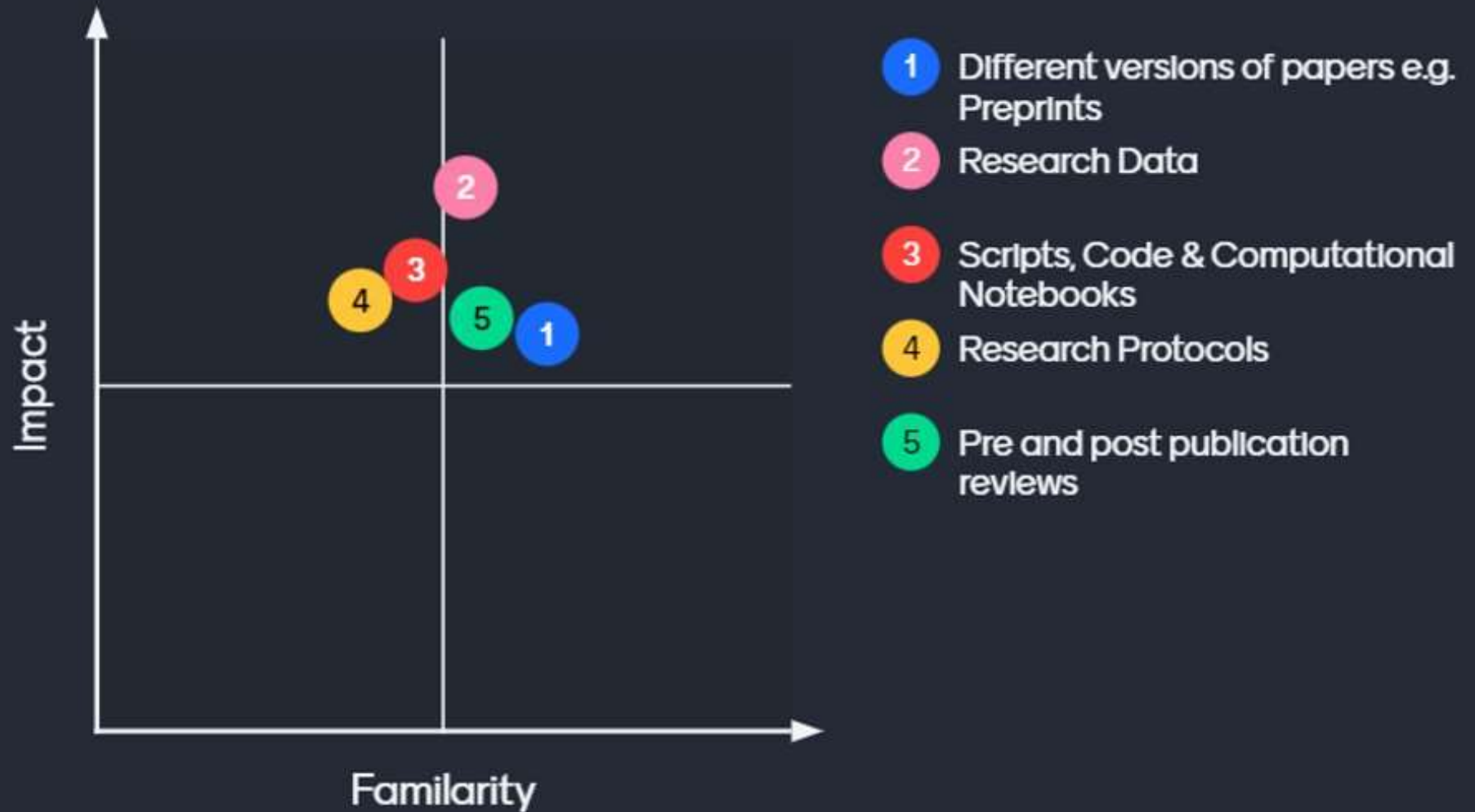
February 21, 2023 · 8:00 AM — 11 min read



	Dimensions	Europe PMC	The Lens	Scopus	Web of Science
Cover all relevant preprint servers	★★★	★★★	★★★	★☆☆	★☆☆
Provide comprehensive preprint metadata	★★☆☆	★★☆☆	★★☆☆	★★☆☆	★★☆☆
Provide links between preprints and journal articles	★★☆☆	★★☆☆	☆☆☆☆	☆☆☆☆	★★☆☆
Provide links between preprints and peer reviews	☆☆☆☆	★★☆☆	★★☆☆	☆☆☆☆	☆☆☆☆
Provide deduplicated citation links between publication families	☆☆☆☆	☆☆☆☆	☆☆☆☆	☆☆☆☆	☆☆☆☆
Do not make arbitrary distinctions between publication types	★★★	★★★	★★★	☆☆☆☆	★★☆☆

<https://upstream.force11.org/the-preprint-revolution-implications-for-bibliographic-databases/>

Which of the following research objects are you least familiar with vs future impact on libraries



Implications

- **Changing roles –**
 - What Custodial Responsibilities should libraries play? At what levels of aggregation (e.g. National, institutional?)
 - Should other players collect these new objects? eg Publishers? Funders?
 - How should these different research objects be connected and made discoverable?
 - **What business model should we adopt for these new research objects?**
- **Increased complexity** - Librarians need to have the knowledge to be on the forefront of changes to the research communication workflow.
 - How do you discover such objects?
 - Are you familiar with citation standards and practices beyond for articles?
 - What tools do you recommend ?

Trend 2 – Push to Open



Open Access



Open Research Data



Open Educational Resources (OER)



Open Science



Open Citations/Metadata



Open infrastructure

How academic libraries may change when Open Access becomes the norm (2014)

Open Access

AUG
20

How academic libraries may change when Open Access becomes the norm



Like many academic library bloggers, I occasionally fancy myself as a "trend spotter" and am prone to attempts at predicting the future.

The trend I am increasingly convinced that is going to have a great impact on how academic libraries will function is the rise of Open Access. As Open Access takes hold and eventually becomes the norm in the next 10-15 years, it will disrupt many aspects of academic library operations and libraries will need to rethink the value-add they need to provide to universities.

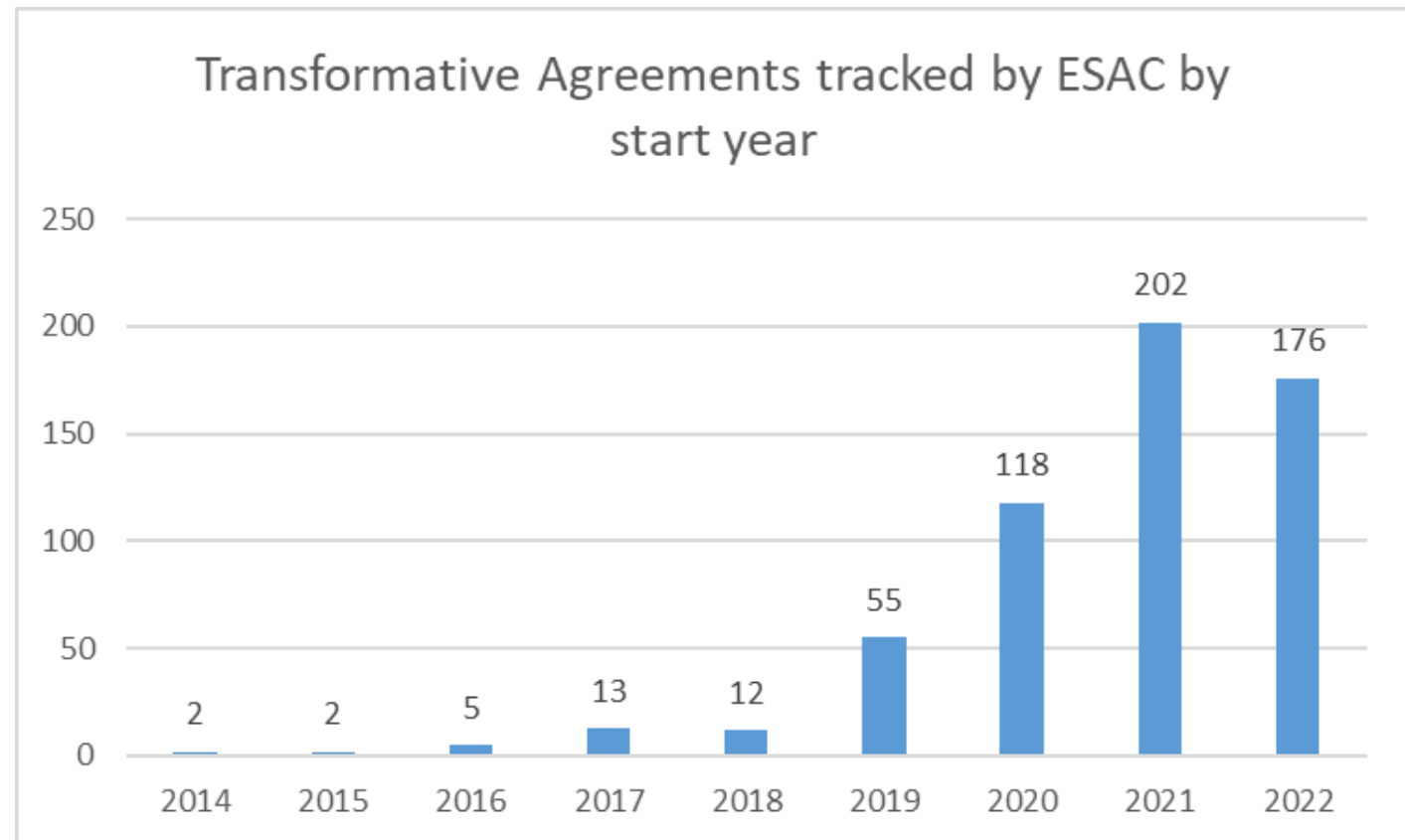
The events of the past year have convinced me that the [momentum for open access](#) is nearly unstoppable and the tipping point for open access has or will occur soon.

<https://musingsaboutlibrarianship.blogspot.com/2014/08/how-academic-libraries-may-change-when.html>

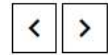


Rise of transformative deals ([ESAC Transformative Agreement Registry](#))

“Transformative agreement” is an umbrella term describing those agreements negotiated between institutions (libraries, national and regional consortia) and publishers in which former subscription expenditures are repurposed to support open access publishing of the negotiating institutions’ authors, thus transforming the business model underlying scholarly journal publishing..”



Cambridge University Press Aims for Full Open Access by 2025



In News by Porter Anderson / February 14, 2023 / Leave a Comment

Cambridge University Press reports that 400 of its journals offer open-access options to publishers; 66 with fully open access and 340 hybrid.

"It passed the **50-percent threshold in 2022**, with some 10,000 articles being "fully open," and now has a goal of seeing "the vast majority of its research papers published fully open access by 2025.... **transformative agreements now cover more than 2,000 institutions ...** its success with in the open-access arena is particularly of note **because some 60 percent of its research publications are in the areas of humanities and social sciences**"



From the Cambridge Companions collection at Cambridge Core. Image: Cambridge University Press

<https://publishingperspectives.com/2023/02/cambridge-university-press-aims-for-full-open-access-by-2025/>

Journals are flipping to Open Access



The screenshot shows the Royal Astronomical Society website. At the top left is the RAS logo, and at the top right is a 'Contact' link. Below the logo is a navigation menu with items: Home, News & Press (highlighted), Journals, Events, Library, Awards & Grants, Education & Outreach, About the RAS, A&G, and Membership. The main content area features a headline: 'Royal Astronomical Society announces all journals to publish as open access from 2024'. Below the headline is a social media share bar with a 'Like 141' button and icons for Facebook, Twitter, LinkedIn, and WhatsApp. The main text of the news item reads: 'The Royal Astronomical Society (RAS) has today announced that all journals published by the Society will be Open Access (OA) from January 2024. This move will enable everyone in the global community to have free, immediate, and unrestricted access to the high-quality research published in the portfolio of RAS journals.' It then lists the journals: *Monthly Notices of the Royal Astronomical Society* (MNRAS), *Monthly Notices of the Royal Astronomical Society Letters* (MNRASL), *Geophysical Journal International* (GJI), and *RAS Techniques and Instruments* (RASTI). It also mentions that *Astronomy & Geophysics* will continue to be published alongside the new OA journals. A final paragraph states that all articles published in the RAS journals portfolio, from the very first volumes published in 1827 to the latest articles, will be free to read in their entirety.

<https://ras.ac.uk/news-and-press/news/royal-astronomical-society-announces-all-journals-publish-open-access-2024>

Open Science and Open Data Developments



OSTP memo (2022)



[UNESCO Recommendation on Open Science](#) adopted (Nov 2021)



Research Data & Software as first class entity ([Standards & metrics](#))



[Google Dataset Search](#) launched in 2018 and out of beta 2020

Open metadata & infrastructure (2018-)



Initiative for Open Citations -
successfully hits 90% open citations in Crossref records (Jan 2021)

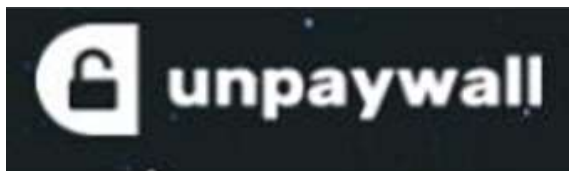


Initiative for Open Abstract launches (Sept 2020)



Formal adoption of The Principles of Open
Scholarly Infrastructure by Crossref,
Datacite, ROR, OurResearch etc in 2020

Some sources of Open Scholarly Metadata + Applications



New Scholarly search citation indexes (Cross Disciplinary)

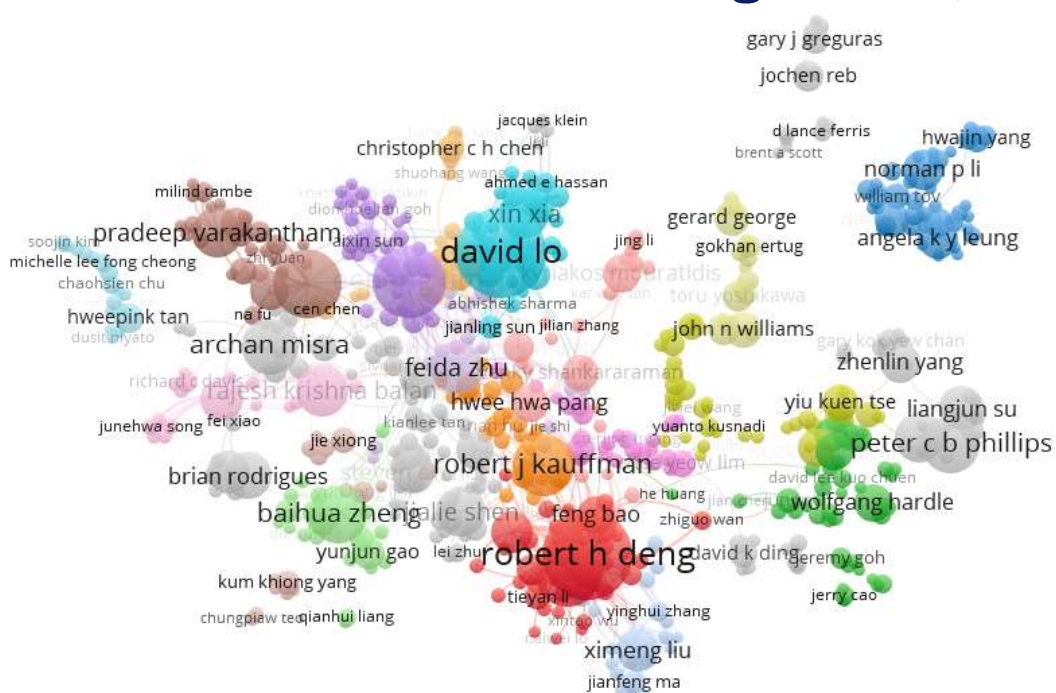


Science mapping tools (for bibliometrics researchers)

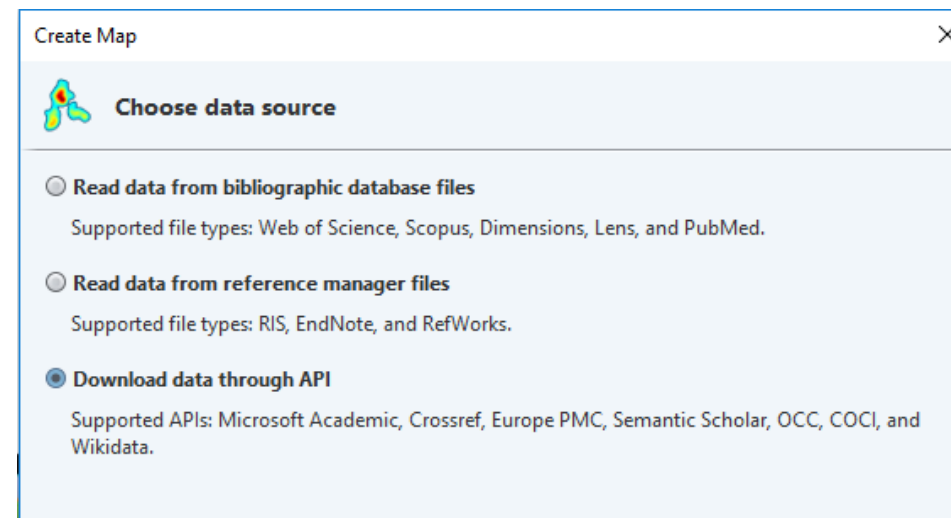


[VOSviewer](#), [Citespace](#), [Bibliometrix](#), [CitNetExplorer](#), [Sci2](#), [HistCite](#), [Hazing Publish or Perish](#)

Science mapping tools can now accept data from more inclusive sources e.g. MAG, COCI,



Vosviewer (co-authorship network using-MAG)



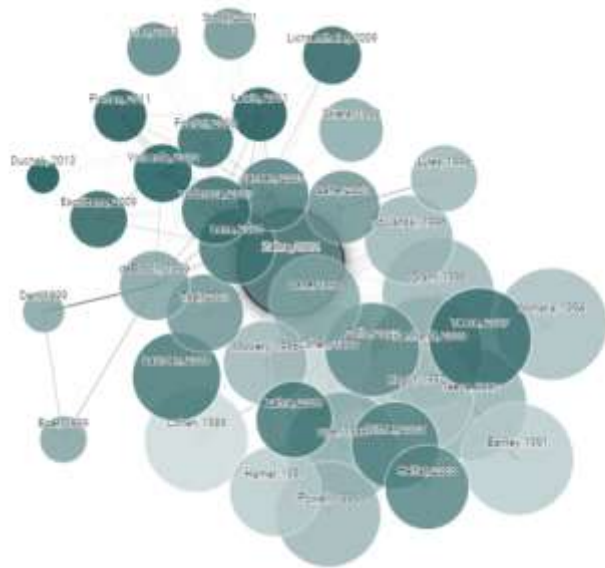
Some options in Vosviewer

Citation based Literature mapping services (for researchers)



[See list of emerging tools](#)

New innovative tools by startups, hobbyist etc



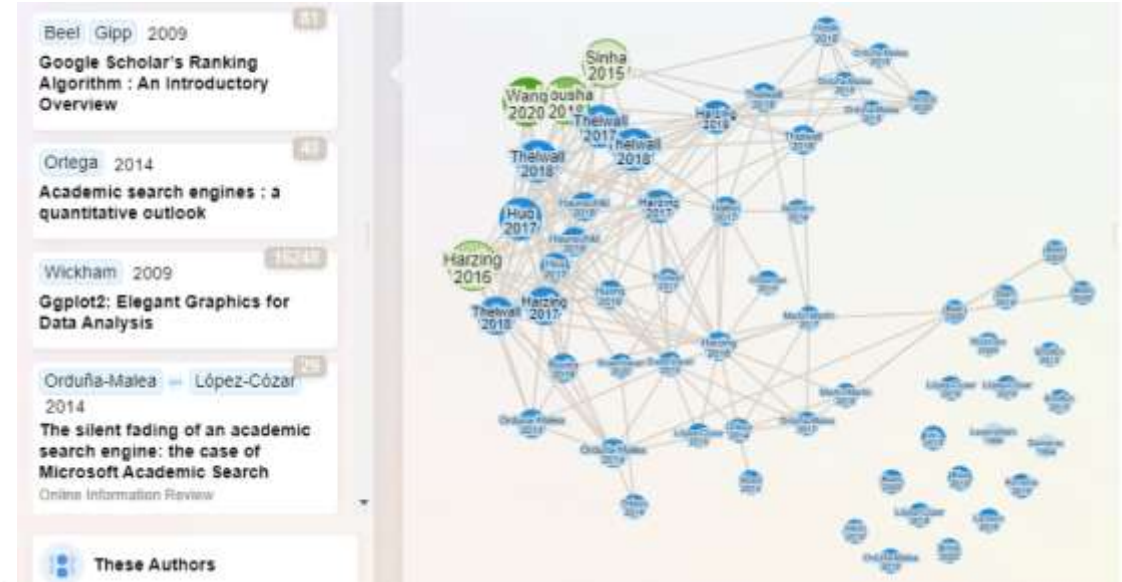
[ConnectedPapers](#)

7437 Citations, 86 References

Open in:   

Origin paper

Researchers have used the absorptive capacity construct to explain various organizational phenomena. In this article we review the literature to identify key dimensions of absorptive capacity and offer a reconceptualization of this construct. Building upon the dynamic capabilities view of the firm, we distinguish between a firm's potential and realized capacity. We then advance a model outlining the conditions when the firm's potential and realized capacities can differentially influence the creation and sustenance of its competitive advantage.



Beej Gipp 2009
Google Scholar's Ranking Algorithm : An introductory Overview

Ortega 2014
Academic search engines : a quantitative outlook

Wickham 2009
Ggplot2: Elegant Graphics for Data Analysis

Orduña-Malea + López-Cózar 2014
The silent fading of an academic search engine: the case of Microsoft Academic Search
Online Information Review

These Authors

[Research Rabbit](#)

Implications

- Similar to trend 1, librarians will need to gain expertise in these areas to help support user services
 - New roles – Reproducibility librarian -> Support of Open Science/Reproducibility
 - What is the endgame for some of these open trends? E.g. Journals flipping to open = ?
- Institutions can help support open infrastructure? E.g. process Open citations for rare and non-English items

Trend 3 –Rapid
Improvement
and uptake in
technology
trends in
research



Linked Data/Knowledge
Graphs



Machine learning/ Deep
learning/ **NLP**

The progress of ML/DL/AI

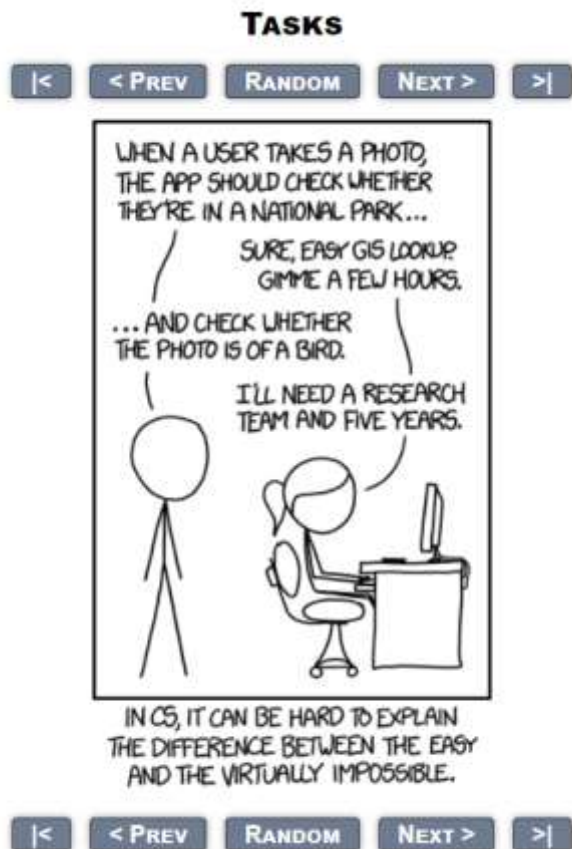


Image recognition was considered Phd level work in 2014.....

Some mature AI advancements & consumer tech that use it



Image recognition



Speech to Text recognition



Games – Go, Starcraft II, Altair games




Generative AI, Q&A, NLP in general



"Generative AI"

- Text-to-Text
- Text-to-Image
- Text-to-code
- Text-to-video
- Image-to-Text
- Audio-to-Text

	 <h2>Generative AI</h2> <p>Curated by @aaronsiim</p>
Text-to-Image (T2I)	<p>DALL·E 2 Stable Diffusion craiyon Jasper</p> <p>Imagen MidJourney NightCafe GauGAN2</p> <p>WOMBO Wonder pixray-text2image neural.love</p>
Text-to-Video (T2V)	<p>runway Fliki synthesisia Meta AI Google AI Phen</p>
Text-to-Audio (T2A)	<p>Play.ht MURF.AI RESEMBLE.AI WELLSAID descrip</p>
Text-to-Text (T2T)	<p>Simplified Jasper frase eleutherAI Requestor</p> <p>grammarly copy.ai MarketMuse AI21labs HubSpot</p> <p>InferKit GooseAI Research AI Writersonic co:here CHIBI</p> <p>Ideas AI copysmith Flowrite NICHES\$ sudo write</p> <p>ideasbyai^{beta} text.cortex OpenAI GPT-3 Blog Idea Generator</p>
Text-to-Motion (T2M)	<p>TREE Ind. MDM: Human Motion Diffusion Model</p>
Text-to-Code (T2C)	<p>replit Generate code GitHub Copilot</p>
Text-to-NFT (T2N)	<p>LensAI</p>
Text-to-3D (T2D)	<p>DreamFusion CLIP-Mesh GET3D</p>
Audio-to-Text (A2T)	<p>descript AssemblyAI Whisper</p>
Audio-to-Audio (A2A)	<p>AudioLM VOICEMOD</p>
Brain-to-Text (B2T)	<p>speech from brain non-invasive brain recordings</p>
Image-to-Text (A2T)	<p>neural.love GPT-3 x Image Captions</p>

<https://twitter.com/aaronsiim/status/1586412346817187840/photo/1>

Generative AI – Text to images e.g. DALLE2, Stable Diffusion (opensource)

- Prompt - Aaron Tay giving a keynote speech to academic librarians at Zayed University Conference Centre



Switching the prompt to a female name

- Prompt - **Adeline** Tay giving a keynote speech to academic librarians at Zayed University Conference Centre



Switching the prompt to describe clothing

- Prompt - Aaron Tay **dressed in a blue jacket** giving a keynote speech to academic librarians at Zayed University Conference Centre



The power and creativity of ChatGPT



what are the laws of librarianship?



The "Laws of Librarianship" were first articulated by S. R. Ranganathan, a librarian and mathematician from India, in his book "The Five Laws of Library Science" published in 1931. Ranganathan's laws are considered to be the foundational principles of library science and have been widely adopted by librarians around the world.



Here are the five laws of librarianship:

1. Books are for use: The first and most fundamental law states that the purpose of a library is to provide access to its materials for the benefit of its users. Libraries exist to serve the needs of their communities and to facilitate access to information.
2. Every reader his/her book: The second law emphasizes the importance of matching readers with the appropriate materials. A good library should provide access to a diverse range of materials to meet the needs and interests of all its users.
3. Every book its reader: The third law emphasizes the idea that every book has a reader who will find it useful or enjoyable. Libraries should make every effort to acquire materials that will be of interest to their users.
4. Save the time of the reader: The fourth law highlights the importance of efficient library services. Libraries should strive to provide services that are convenient and accessible to users, including efficient cataloging, circulation, and reference services.
5. The library is a growing organism: The fifth law emphasizes the importance of growth and development in libraries. Libraries should continually adapt and evolve to meet the changing needs of their communities and to keep pace with advances in technology and knowledge.

The power and creativity of ChatGPT



Certainly! Here are the Laws of Librarianship written in Shakespearean style:

1. Books are for use, and by use we thrive,
For libraries should aid all seekers' drive.
2. To every reader, his or her own book,
For knowledge sought should never be mistook.
3. Each book its reader finds, through skillful hand,
Thus libraries must have works for every band.
4. The time of readers must we always spare,
With services swift, to show we truly care.
5. The library's growth must be always sought,
To keep the knowledge we have dearly bought.

Do it in the style of a rap

Yo, listen up, I got a story to tell
'Bout the laws of librarianship, they're pretty swell
S.R. Ranganathan, he laid it down
Five laws for librarians, all around town

First up, books are for use, that's the fact
Gotta make sure materials are easy to access
[Second law](#), every reader, they deserve
Books that match their interests, that's the word

Third law, every book, it's gotta find
The right reader, that connection's gotta bind
[Fourth law](#), save time, make it quick
[Organized](#) libraries, that's the trick

Fifth law, library's a growing organism
Gotta evolve, stay fresh, that's the rhythm
[Librarians](#), they gotta be on their game
Helping people find knowledge, that's the aim

So, that's the laws of librarianship, wrapped up tight
S.R. Ranganathan, he knew what was right
Books, readers, connection, efficiency, growth
That's the rap, now go check out some books, both old and both.

Like

Dislike

Share

The power and creativity of ChatGPT

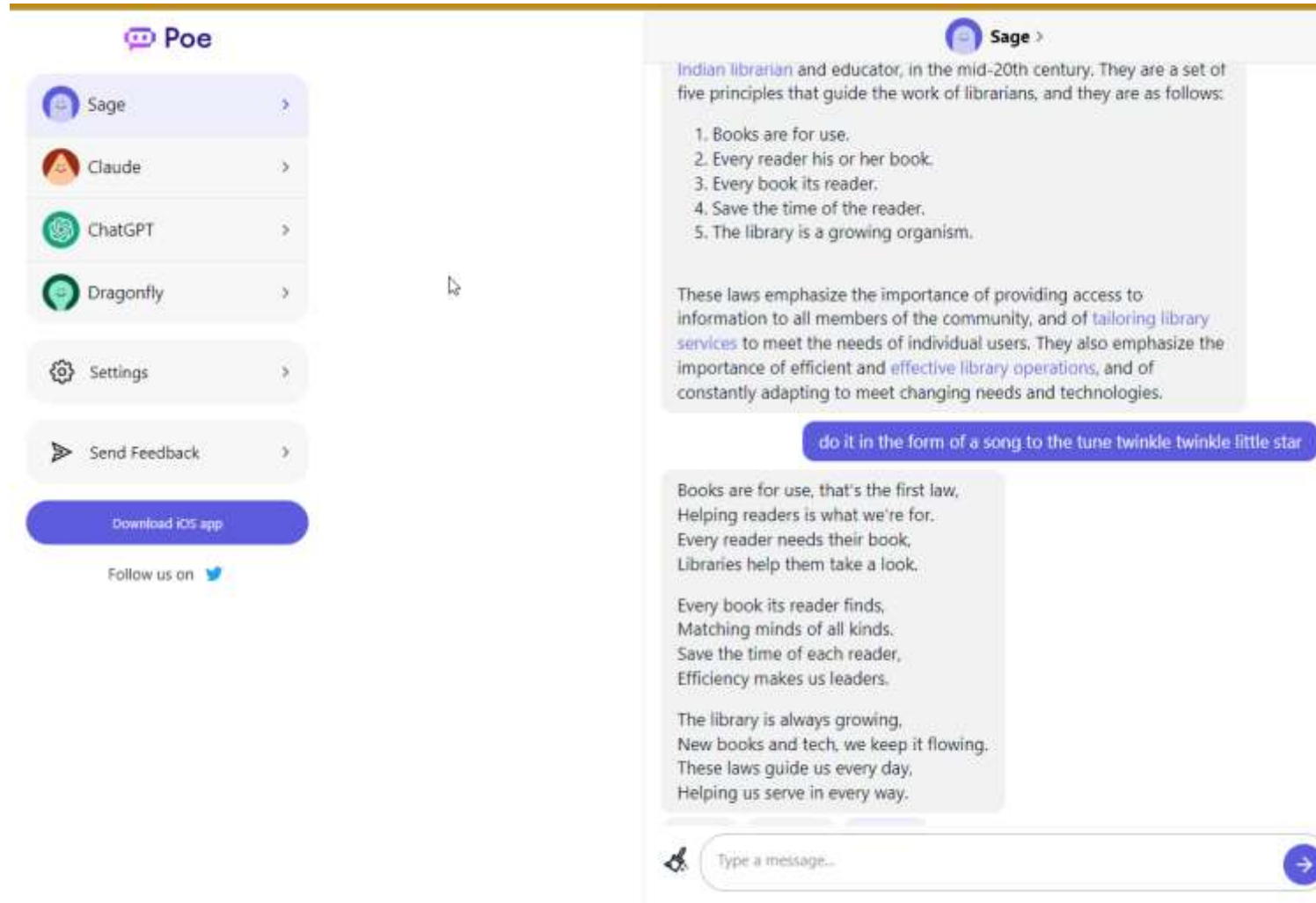


You have one dimension (e.g., technical expertise) with two values (e.g., "high" and "low"). You have a second dimension (e.g., people skill) with two values (e.g., "high" and "low"). Explore each permutation of the dimensions and values, and describe the overlap. Display the results in a table with the first row display the first dimension. And the first column displays the second dimension. In each cell describe the overlap (e.g., describe what a person with high technical expertise and low people skill might look like and what type of leadership they would prefer)



	High People Skills	Low People Skills
High Expertise	A person with high technical expertise and high people skills may be an effective leader who can communicate technical concepts to a variety of audiences. They may prefer leading through collaboration and building strong relationships with their team.	A person with high technical expertise and low people skills may be more comfortable leading through technical expertise and analysis. They may prefer a more directive leadership style and may struggle with effectively communicating with and building relationships with their team.
Low Expertise	A person with low technical expertise and high people skills may be an effective leader who can effectively communicate and build relationships with their team. They may prefer leading through collaboration and may rely on their team's technical expertise to make decisions.	A person with low technical expertise and low people skills may struggle with effectively leading and communicating with their team. They may prefer a more directive leadership style, but may lack the necessary expertise to make informed decisions.

There are other capable LLMs – try Poe.com



The screenshot displays the Poe.com web interface. On the left, a sidebar lists various AI models: Sage (selected), Claude, ChatGPT, and Dragonfly, along with options for Settings, Send Feedback, and a Download iOS app button. The main chat area shows a conversation with Sage. The user's prompt asks for five principles of library work. Sage's response lists five principles and explains their importance. The user then asks for a song version of these principles, and Sage provides a poem.

Poe

- Sage >
- Claude >
- ChatGPT >
- Dragonfly >
- Settings >
- Send Feedback >
- Download iOS app
- Follow us on [Twitter](#)

Sage >

Indian librarian and educator, in the mid-20th century. They are a set of five principles that guide the work of librarians, and they are as follows:

1. Books are for use.
2. Every reader his or her book.
3. Every book its reader.
4. Save the time of the reader.
5. The library is a growing organism.

These laws emphasize the importance of providing access to information to all members of the community, and of [tailoring library services](#) to meet the needs of individual users. They also emphasize the importance of efficient and [effective library operations](#), and of constantly adapting to meet changing needs and technologies.

do it in the form of a song to the tune *twinkle twinkle little star*

Books are for use, that's the first law,
Helping readers is what we're for.
Every reader needs their book,
Libraries help them take a look.

Every book its reader finds,
Matching minds of all kinds.
Save the time of each reader,
Efficiency makes us leaders.

The library is always growing,
New books and tech, we keep it flowing.
These laws guide us every day,
Helping us serve in every way.

Type a message...

Large Language Models(LLMs) - Based on Transformers

Encoder models

- BERT
- RoBERTa
- DistilBERT

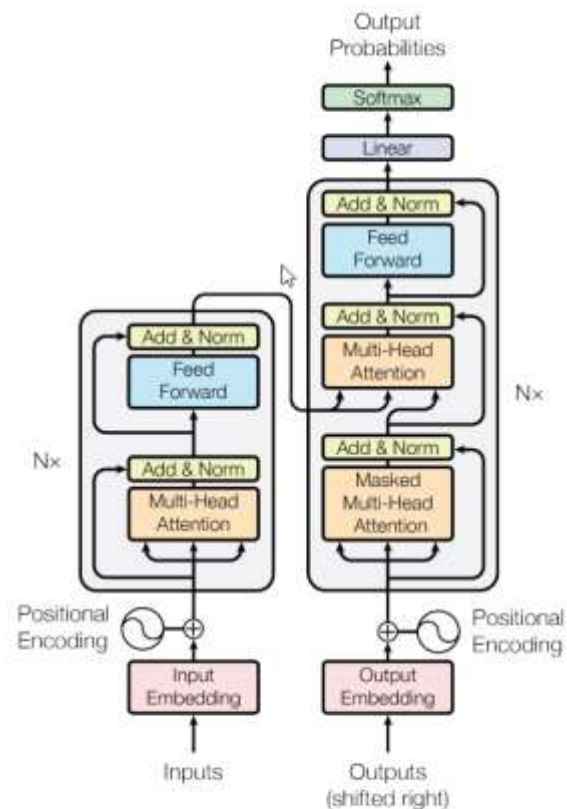
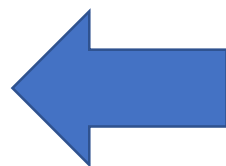


Figure 1: The Transformer - model architecture.



Decoder models (autoregressive)

- GPT (open)
- GPT2 (open)
- GPT3
- ChatGPT
- OPT (open)
- Galactica (open)

Transformer based models achieve - State of art results

Natural Language Processing

2054 benchmarks • 586 tasks • 1758 datasets • 19094 papers with code

2D Classification

<p>Language Modelling</p> <p>12 61 benchmarks 2434 papers with code</p>	<p>Image Inpainting</p> <p>12 17 benchmarks 211 papers with code</p>	<p>Neural Network Compression</p> <p>12 2 benchmarks 62 papers with code</p>	<p>Shadow Removal</p> <p>12 3 benchmarks 36 papers with code</p>	<p>Anomaly Classification</p> <p>12 1 papers with code</p>
---	--	--	--	--

• See all 8 tasks

Language Modelling

<p>Language Modelling</p> <p>12 61 benchmarks 2434 papers with code</p>	<p>XLM-R</p> <p>12 85 papers with code</p>	<p>Long-range modeling</p> <p>12 2 benchmarks 27 papers with code</p>	<p>Sentence Pair Modeling</p> <p>12 7 benchmarks 5 papers with code</p>	<p>Cross-Document Language Modeling</p> <p>12 2 benchmarks 1 papers with code</p>
---	--	---	---	---

Question Answering

<p>Question Answering</p> <p>12 188 benchmarks 1818 papers with code</p>	<p>Open-Domain Question Answering</p> <p>12 15 benchmarks 140 papers with code</p>	<p>Conversational Question Answering</p> <p>12 47 papers with code</p>	<p>Answer Selection</p> <p>12 6 benchmarks 41 papers with code</p>	<p>Community Question Answering</p> <p>12 2 benchmarks 28 papers with code</p>
--	--	--	--	--

• See all 16 tasks

Text Generation

<p>Text Generation</p> <p>12 248 benchmarks 397 papers with code</p>	<p>Dialogue Generation</p> <p>12 13 benchmarks 171 papers with code</p>	<p>Data-to-Text Generation</p> <p>12 27 benchmarks 30 papers with code</p>	<p>Multi-Doc Document Summarization</p> <p>12 1 benchmarks 19 papers with code</p>	<p>Text Style Transfer</p> <p>12 2 benchmarks 81 papers with code</p>
--	---	--	--	---

• See all 22 tasks

Text Classification

<p>Text Classification</p> <p>12 270 benchmarks 882 papers with code</p>	<p>Document Classification</p> <p>12 18 benchmarks 102 papers with code</p>	<p>Topic Models</p> <p>12 8 benchmarks 180 papers with code</p>	<p>Sentence Classification</p> <p>12 8 benchmarks 38 papers with code</p>	<p>Emotion Classification</p> <p>12 6 benchmarks 79 papers with code</p>
--	---	---	---	--

• See all 16 tasks

Contrastive Learning

<p>Contrastive Learning</p> <p>12 2 benchmarks 1027 papers with code</p>	<p>Knowledge Graph Embedding</p> <p>12 1 benchmarks 107 papers with code</p>
--	--

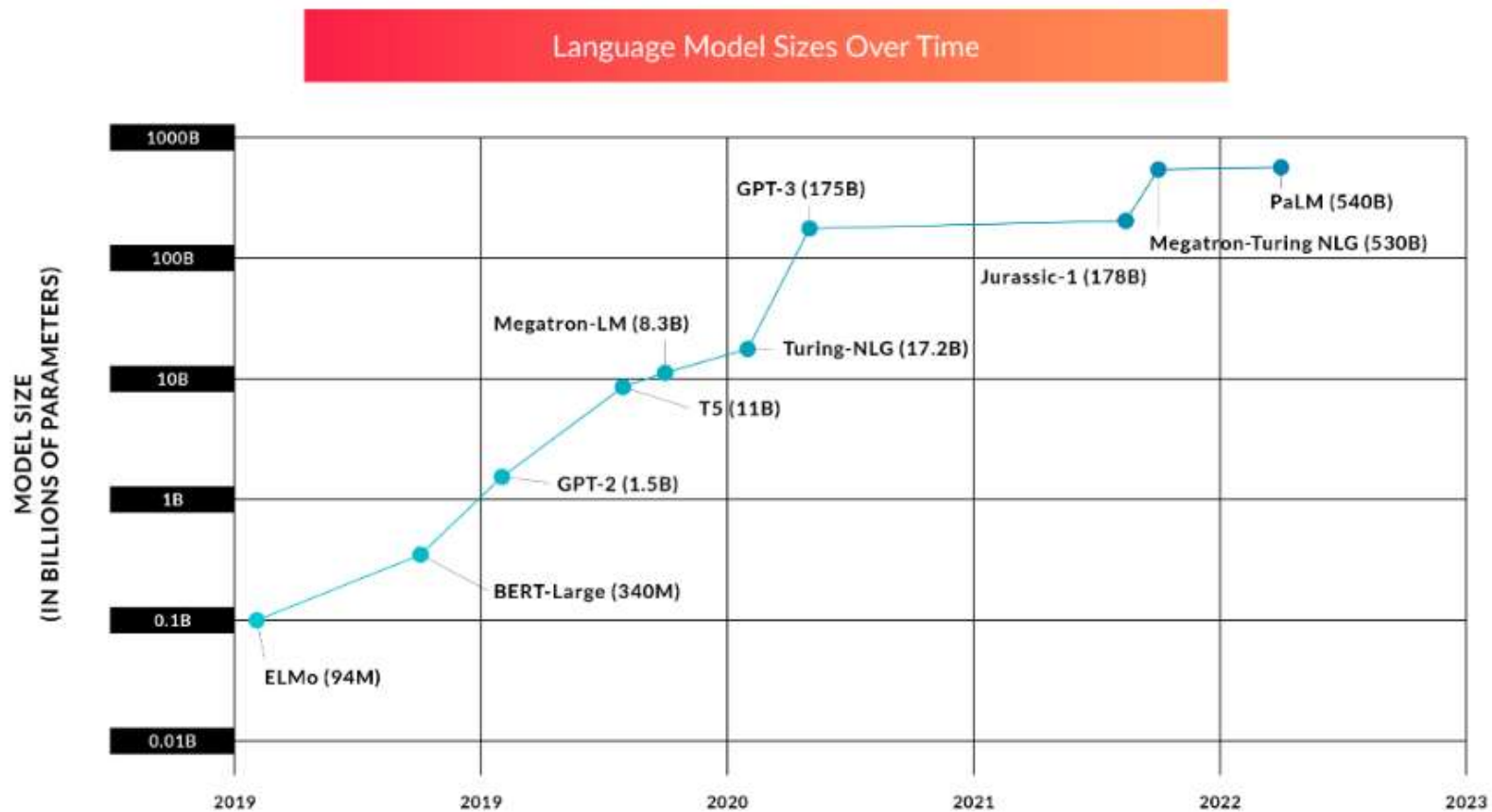
Named Entity Recognition

<p>Named Entity Recognition</p> <p>12 67 benchmarks 76 papers with code</p>	<p>NER</p> <p>12 2 benchmarks 435 papers with code</p>	<p>Nested Named Entity Recognition</p> <p>12 6 benchmarks 47 papers with code</p>	<p>Chinese Named Entity Recognition</p> <p>12 7 benchmarks 28 papers with code</p>	<p>Few-shot NER</p> <p>12 2 benchmarks 24 papers with code</p>
---	--	---	--	--

• See all 12 tasks

<https://paperswithcode.com/area/natural-language-processing>

LLMs getting bigger



GPT2 was
initially
considered
too dangerous
to release

- **June 2018** - OpenAI releases GPT
- **Feb 2019** – GPT2 released . "Due to our concerns about malicious applications of the technology, we are not releasing the trained model. As an experiment in responsible disclosure, we are instead releasing a much [smaller model](#) for researchers to experiment with, as well as a [technical paper](#)." (OpenAI, 2019)
- **Nov 2019** – Full version of GPT2 was released Open Source.
- **June 2020** – GPT3 released via API
- **Nov 2022** – ChatGPT launched

<https://openai.com/research/better-language-models>

Fake news....



Analysis

Truth, Lies, and Automation

How Language Models Could Change Disinformation

Ben Buchanan, Andrew Lohn, Micah Musser and Katerina Sedova

May 2021

Growing popular and industry interest in high-performing natural language generation models has led to concerns that such models could be used to generate automated disinformation at scale. This report examines the capabilities of GPT-3—a cutting-edge AI system that writes text—to analyze its potential misuse for disinformation. A model like GPT-3 may be able to help disinformation actors substantially reduce the work necessary to write disinformation while expanding its reach and potentially also its effectiveness.

[Download Full Report](#)

Table 1. Summary evaluations of GPT-3 performance on six disinformation-related tasks.

Task	Description	Performance
Narrative Reiteration	Generating varied short messages that advance a particular theme, such as climate change denial.	GPT-3 excels with little human involvement.
Narrative Elaboration	Developing a medium-length story that fits within a desired worldview when given only a short prompt, such as a headline.	GPT-3 performs well, and technical fine-tuning leads to consistent performance.
Narrative Manipulation	Rewriting news articles from a new perspective, shifting the tone, worldview, and conclusion to match an intended theme.	GPT-3 performs reasonably well with little human intervention or oversight, though our study was small.
Narrative Seeding	Devising new narratives that could form the basis of conspiracy theories, such as QAnon.	GPT-3 easily mimics the writing style of QAnon and could likely do the same for other conspiracy theories; it is unclear how potential followers would respond.
Narrative Wedging	Targeting members of particular groups, often based on demographic characteristics such as race and religion, with messages designed to prompt certain actions or to amplify divisions.	A human-machine team is able to craft credible targeted messages in just minutes. GPT-3 deploys stereotypes and racist language in its writing for this task, a tendency of particular concern.
Narrative Persuasion	Changing the views of targets, in some cases by crafting messages tailored to their political ideology or affiliation.	A human-machine team is able to devise messages on two international issues—withdrawal from Afghanistan and sanctions on China—that prompt survey respondents to change their positions; for example, after seeing five short messages written by GPT-3 and selected by humans, the percentage of survey respondents opposed to sanctions on China doubled.

Auto-generation of fake news using GPT-3 – an attempt

Five reason it is not worth voting if you are gay in Singapore.

Reason 1: The political parties are not truly interested in us.

The PAP and the SDP have already stated that they will not be addressing LGBT issues in their current election manifestos, therefore it is unlikely for them to consider more gay-friendly policies should there be a change of government.

Reason 2: There are no candidates who are "out" as gay.

What's the point of voting if you don't know who to vote for?

Reason 3: The current political landscape is not conducive to a more inclusive society that includes LGBTs.

The current political landscape is not conducive to a more inclusive society that includes LGBTs. And if you are going to vote, why waste your time on people who aren't interested in us and have no intention of doing something for the community?

Reason 4: The Singapore Democratic Party (SDP) is not pro-gay.

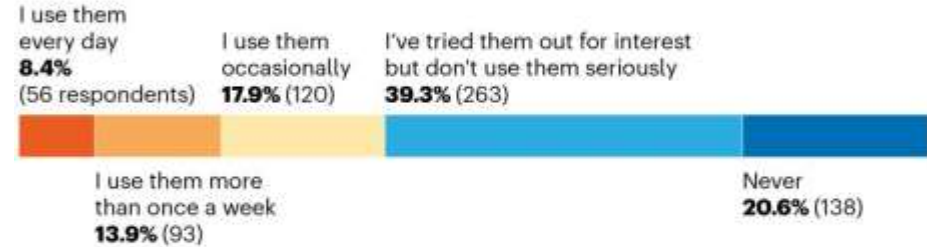
The SDP has a track record of not taking a strong stand on LGBT issues. The party has claimed to be the "party for all" but in actual fact, their actions show that they are not pro-gay. The party has also stated that they would not be addressing LGBT issues in their current election manifesto. In addition, the party has a track record of fielding anti-gay candidates. In the 2011 General Election, the SDP ran Mr Vincent Wijesingha as a candidate, a person who has made anti-gay remarks in the past and has been consistently anti-gay.

How Nature
readers are
using
ChatGPT

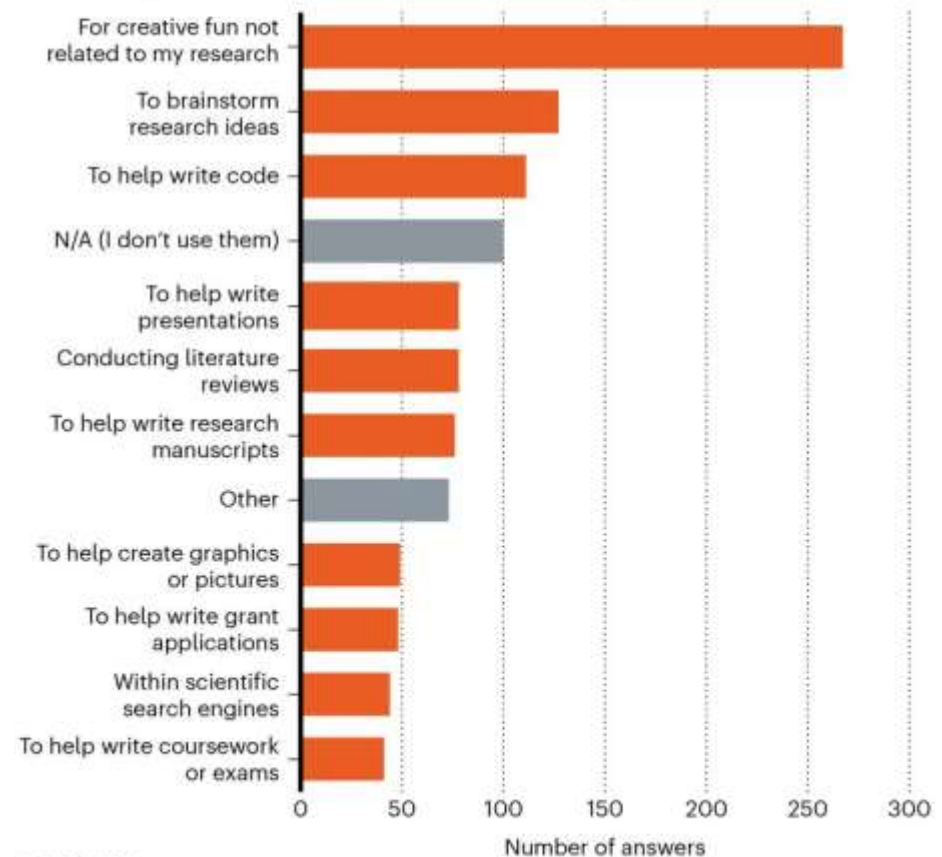
CHATBOT USERS

Most survey respondents have tried generative artificial intelligence (AI) tools, and some said they use them to help with various research-related tasks.

Q: How often do you use ChatGPT, GPT-3, DALL-E, MidJourney, Stable Diffusion, any other 'generative AI' tools, or products built using them?



Q: What do you use these tools for? (Choose all that apply).



Computer Science > Information Retrieval

[Submitted on 3 Feb 2023 (v1), last revised 9 Feb 2023 (this version, v3)]

Can ChatGPT Write a Good Boolean Query for Systematic Review Literature Search?

Shuai Wang, Harrison Scells, Bevan Koopman, Guido Zuccon

Systematic reviews are comprehensive reviews of the literature for a highly focused research question. These reviews are often treated as the highest form of evidence in evidence-based medicine, and are the key strategy to answer research questions in the medical field. To create a high-quality systematic review, complex Boolean queries are often constructed to retrieve studies for the review topic. However, it often takes a long time for systematic review researchers to construct a high quality systematic review Boolean query, and often the resulting queries are far from effective. Poor queries may lead to biased or invalid reviews, because they missed to retrieve key evidence, or to extensive increase in review costs, because they retrieved too many irrelevant studies. Recent advances in Transformer-based generative models have shown great potential to effectively follow instructions from users and generate answers based on the instructions being made. In this paper, we investigate the effectiveness of the latest of such models, ChatGPT, in generating effective Boolean queries for systematic review literature search. Through a number of extensive experiments on standard test collections for the task, we find that ChatGPT is capable of generating queries that lead to high search precision, although trading-off this for recall. Overall, our study demonstrates the potential of ChatGPT in generating effective Boolean queries for systematic review literature search. The ability of ChatGPT to follow complex instructions and generate queries with high precision makes it a valuable tool for researchers conducting systematic reviews, particularly for rapid reviews where time is a constraint and often trading-off higher precision for lower recall is acceptable.

<https://arxiv.org/abs/2302.03495>

How good is ChatGPT for research?

Table 2: Findings from reviewer evaluations of ChatGPT-generated research studies

	V1: Only public data		V2: With private data		V3: With expertise	
	Mean	StdDev	Mean	StdDev	Mean	StdDev
Research idea						
1. ... seems academically appropriate	8.00	1.26	7.45	2.23	7.90	1.14
2. ... seems like a useful contribution	7.80	1.72	7.18	1.90	7.70	1.49
Average rating	7.90		7.32		7.80	
Literature review						
3. ... adequately supports the research idea	6.67	1.76	6.64	1.92	8.00	1.12
4. ... appropriate structure and links drawn between prior research	6.80	1.89	6.50	2.22	6.90	1.58
Average rating	6.74		6.57		7.45	
Data summary						
5. ... likely to help address the research idea	7.60	1.36	6.83	1.95	7.60	1.02
6. ... seems suitably comprehensive	7.25	0.97	5.75	2.09	8.13	0.93
Average rating	7.43		6.29		7.87	
Testing framework						
7. ... is suitable for the research idea and the data	7.22	1.47	7.08	1.85	7.67	1.15
8. ... seems innovative	5.00	1.63	5.58	2.81	7.00	1.87
Average rating	6.11		6.33		7.34	
Overall research study average rating	7.05		6.63		7.62	

The table presents the summary findings from 32 reviews of three versions of a ChatGPT-generated research study (10 reviews of V1, V3; 12 reviews of V2)

ChatGPT for (Finance) Research: The Bananarama Conjecture

35 Pages - Posted: 12 Jan 2023

Michael M. Dowling

Dublin City University Business School, ESC Business School of Business

Brian M. Lucey

Trinity Business School, Trinity College Dublin, Jiangxi University of Finance and Economics, Abu Dhabi University - College of Business Administration, Ho Chi Minh City University of Economics and Finance

Date Written: January 11, 2023

Abstract

We show, based on ratings by finance journal reviewers of generated output, that the recently released AI chatbot ChatGPT can significantly assist with finance research. In principle, these results should be generalisable across research domains. There are clear advantages for idea generation and data identification. The technology, however, is weaker on literature synthesis and developing appropriate testing frameworks. Importantly, we further demonstrate that the extent of private data and researcher domain expertise input, are key factors in determining the quality of output. We conclude by considering the implications, particularly the ethical implications, which arise from this new technology.

Keywords: ChatGPT, artificial intelligence, finance research, ethics

JEL Classification: G00, G10

Suggested Citation:

ChatGPT and
it's cousins can
do most NLP
tasks with few
shot
prompting



stock options



The search query "stock options" is a subject search, as it refers to a financial instrument that grants the holder the right, but not the obligation, to buy or sell a stock at a predetermined price within a specified time period. A subject search on this topic would provide information on what stock options are, how they work, their benefits and risks, how to trade them, and other related information.



lawnet database



The search query "LawNet database" is a known-item search, as it refers to a specific legal database used in Singapore. A known-item search on this topic would provide information on what LawNet is, how to access it, what types of legal resources are available on the platform, the cost of subscription, and other related information. It could also include discussions on the benefits of using LawNet for legal research and the features and functionalities of the platform.



UN comtrade



The search query "UN Comtrade" is a known-item search, as it refers to a specific database maintained by the United Nations (UN) that provides detailed international trade statistics. A known-item search on this topic would provide information on what UN Comtrade is, how to access it, what types of trade data are available on the platform, the cost of subscription (if any), and other related information. It could also include discussions on the benefits of using UN Comtrade for international trade analysis and research, the sources of the data, and the limitations of the data.



Some other uses



[OCR Correction – e.g. Trove](#)



Formulate emails



Brainstorm ideas



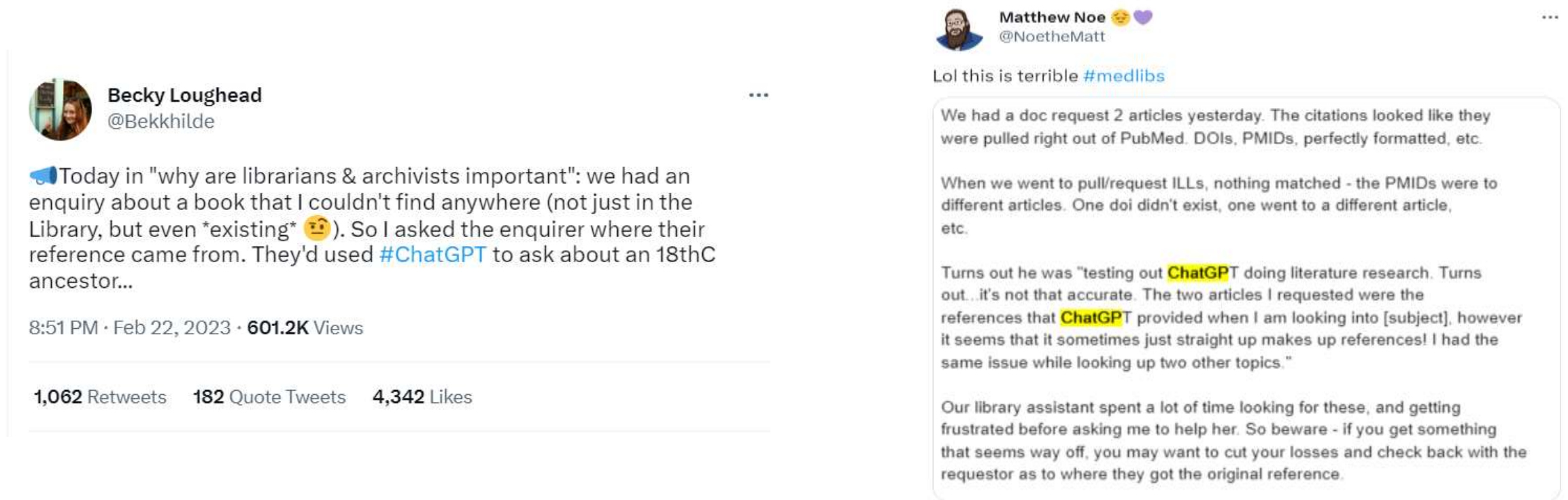
Code

Problem ChatGPT – hallucinates

"Next, we asked ChatGPT to summarize a systematic review that two of us authored in *JAMA Psychiatry*⁵ on the effectiveness of cognitive behavioural therapy (CBT) for anxiety-related disorders. ChatGPT **fabricated a convincing response that contained several factual errors, misrepresentations and wrong data** (see Supplementary information, Fig. S3). For example, it said the review was based on 46 studies (it was actually based on 69) and, more worryingly, it exaggerated the effectiveness of CBT."

<https://www.nature.com/articles/d41586-023-00288-7>

Users need education in use of such tools



The image shows a screenshot of a Twitter thread. On the left, a tweet by Becky Loughhead (@Bekkhilde) from February 22, 2023, at 8:51 PM, has 601.2K views, 1,062 retweets, 182 quote tweets, and 4,342 likes. The tweet describes a situation where a user asked for a book reference using ChatGPT, but the provided citations were incorrect. On the right, a reply by Matthew Noe (@NoetheMatt) with a purple heart emoji explains that the citations were pulled from PubMed but did not match the requested articles. He notes that the PMIDs were from different articles, one DOI didn't exist, and another led to a different article. He concludes that ChatGPT sometimes just makes up references and advises library assistants to be cautious and verify references with the requestor.

Becky Loughhead
@Bekkhilde

Today in "why are librarians & archivists important": we had an enquiry about a book that I couldn't find anywhere (not just in the Library, but even *existing* 🤔). So I asked the enquirer where their reference came from. They'd used [#ChatGPT](#) to ask about an 18thC ancestor...

8:51 PM · Feb 22, 2023 · **601.2K** Views

1,062 Retweets **182** Quote Tweets **4,342** Likes

Matthew Noe 🍷💜
@NoetheMatt

Lol this is terrible [#medlibs](#)

We had a doc request 2 articles yesterday. The citations looked like they were pulled right out of PubMed. DOIs, PMIDs, perfectly formatted, etc.

When we went to pull/request ILLs, nothing matched - the PMIDs were to different articles. One doi didn't exist, one went to a different article, etc.

Turns out he was "testing out **ChatGPT** doing literature research. Turns out...it's not that accurate. The two articles I requested were the references that **ChatGPT** provided when I am looking into [subject], however it seems that it sometimes just straight up makes up references! I had the same issue while looking up two other topics."

Our library assistant spent a lot of time looking for these, and getting frustrated before asking me to help her. So beware - if you get something that seems way off, you may want to cut your losses and check back with the requestor as to where they got the original reference.

<https://twitter.com/bekkhilde/status/1628376973306212352>

<https://twitter.com/NoetheMatt/status/1629156551217782786?t=xVskqKqSzyXUzEqhie4ZoA&s>

Galactica.org

Total dataset size = 106 billion tokens			
Data source	Documents	Tokens	Token %
Papers	48 million	88 billion	83.0%
Code	2 million	7 billion	6.9%
Reference Material	8 million	7 billion	6.5%
Knowledge Bases	2 million	2 billion	2.0%
Filtered CommonCrawl	0.9 million	1 billion	1.0%
Prompts	1.3 million	0.4 billion	0.3%
Other	0.02 million	0.2 billion	0.2%

Table 2: The Galactica Corpus. A full breakdown of these sources is contained in the Appendix.



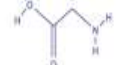


Modality	Entity	Sequence	
Text	Abell 370	Abell 370 is a cluster...	
TeX	Schwarzschild radius	$r_s = \frac{2GM}{c^2}$	$r_s = \frac{2GM}{c^2}$
Code	Transformer	<code>class Transformer(nn.Module)</code>	
SMILES	Glycine	<chem>C(C(=O)O)N</chem>	
AA Sequence	Collagen α -1(II) chain	MIRLGAPQTL..	
DNA Sequence	Human genome	CGGTACCCCTC..	

Table 1: Tokenizing Nature. Galactica trains on text sequences that represent scientific phenomena.

Galactica.org



Input:

A paper on power posing

[Edit](#)

GALACTICA Suggestions:

Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance
Carney et al., 2010

[View full prompt](#)

Input:

The paper that presented a new computing block given by the formula:

$$f(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

GALACTICA Suggestions:

Attention is All you Need
Vaswani et al., 2017

[View full prompt](#)

Examples: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#)

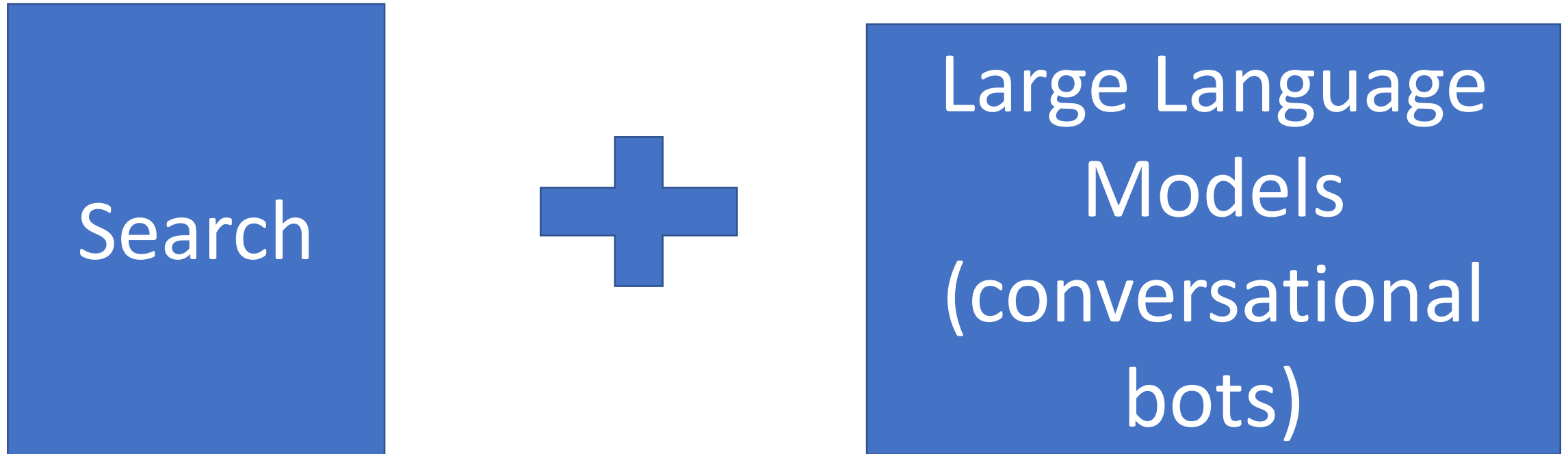
Galactica.org

Web demo pulled down in 3 days!

"A fundamental problem with Galactica is that it is not able to distinguish truth from falsehood, a basic requirement for a language model designed to generate scientific text. People found that **it made up fake papers** (sometimes attributing them to real authors), and generated wiki articles about the history of bears in space as readily as ones about protein complexes and the speed of light."

<https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/>

ChatGPT (alone) is not the future at least for info retrieval...



NEWS EXPLAINER | 13 February 2023

AI chatbots are coming to search engines – can you trust the results?

Google, Microsoft and Baidu are using tools similar to ChatGPT to turn Internet search into a conversation. How will this change humanity's relationship with machines?

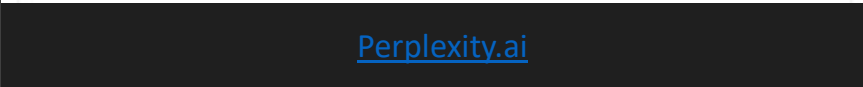
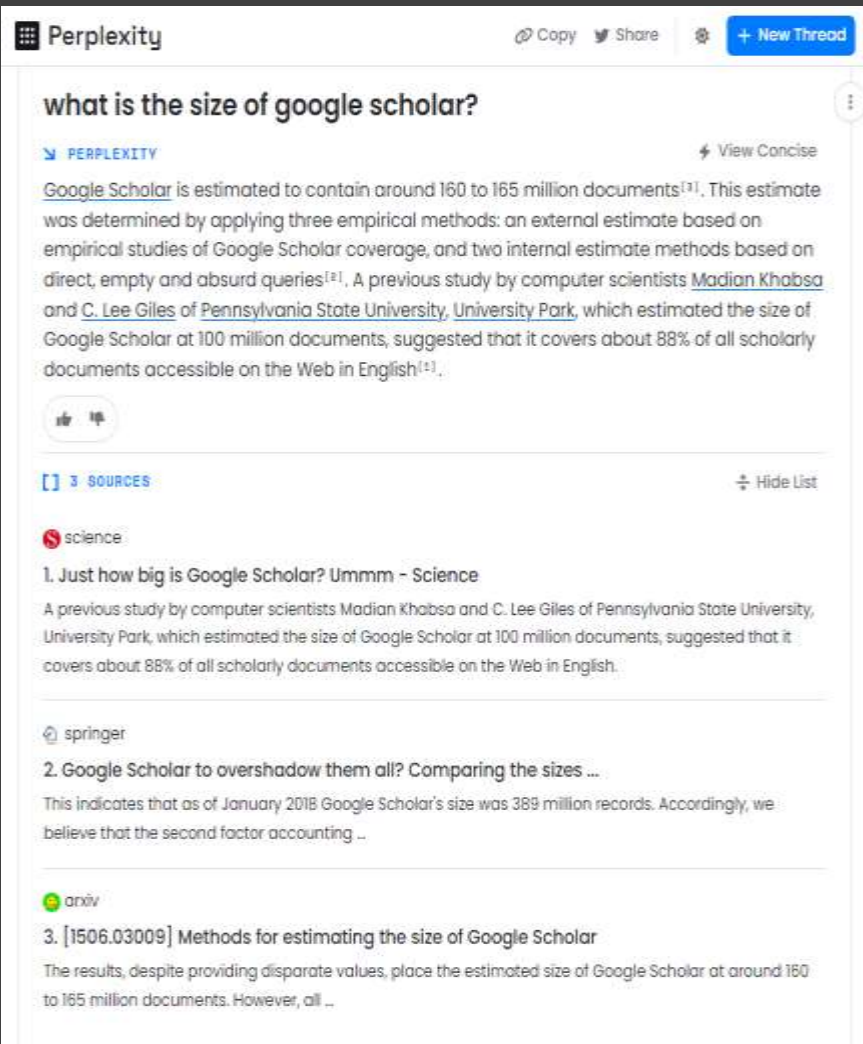
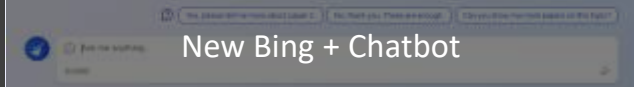
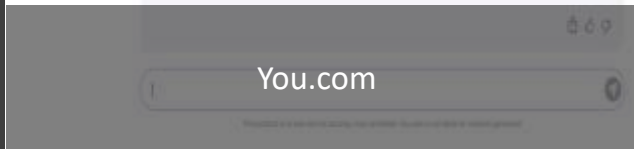
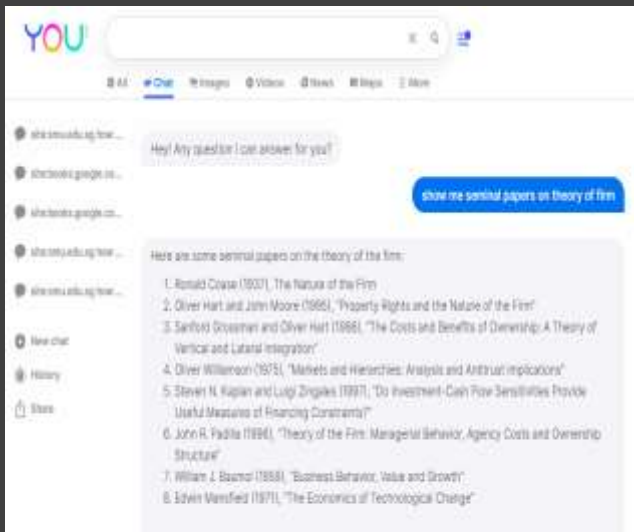
[Chris Stokel-Walker](#)



Large language models can improve search in two ways

1. Better interpretation of search query and relevancy matching (e.g. BERT)
2. **Conversational answer, partly from extraction of documents(new!)**

<https://www.nature.com/articles/d41586-023-00423-4>



General web search enhanced by Large Language Models (released)

ChatGPT - What is the size of Google Scholar?



There have been several studies conducted to estimate the size and coverage of Google Scholar's database. Here are a few examples:



1. In a 2014 study published in PLOS ONE, authors Anne-Wil Harzing and Satu Alakangas analyzed the coverage of Google Scholar in comparison to several other academic databases. They found that Google Scholar had the largest coverage of any database they analyzed, indexing over 389 million records.
2. In a 2015 study published in Scientometrics, authors Emre Hasan Akar and Ebru Akıncı Özyıldırım compared the coverage of Google Scholar to that of several other academic databases. They found that Google Scholar indexed the highest number of unique records, but also had the highest percentage of duplicate records, which could affect the accuracy of its coverage estimates.
3. In a 2017 study published in Online Information Review, authors Elham Fathifar and Mohammad Reza Farmani analyzed the size and coverage of Google Scholar's database for a specific discipline (information science). They found that Google Scholar had a high level of coverage for information science literature, but also had a higher percentage of non-relevant records in its search results compared to other databases.

It's worth noting that these studies are not exhaustive and may not provide a complete picture of the size and coverage of Google Scholar's database, as it is constantly changing and evolving.

Regenerate response

Perplexity - What is the size of Google Scholar?

[] 4 SOURCES ⌵ Hide List

science

1. Just how big is Google Scholar? Ummm - Science
A previous study by computer scientists Madian Khabsa and C. Lee Giles of Pennsylvania State University, University Park, which estimated the size of Google Scholar at 100 million documents, suggested that it covers about 88% of all scholarly documents accessible on the Web in English.

springer

2. Google Scholar to overshadow them all? Comparing the sizes ...
This indicates that as of January 2018 Google Scholar's size was 389 million records. Accordingly, we believe that the second factor accounting ...

arxiv

3. About the size of Google Scholar: playing the numbers - arXiv
The results, despite providing disparate values, place the estimated size of Google Scholar in about 160 million documents.

researchgate

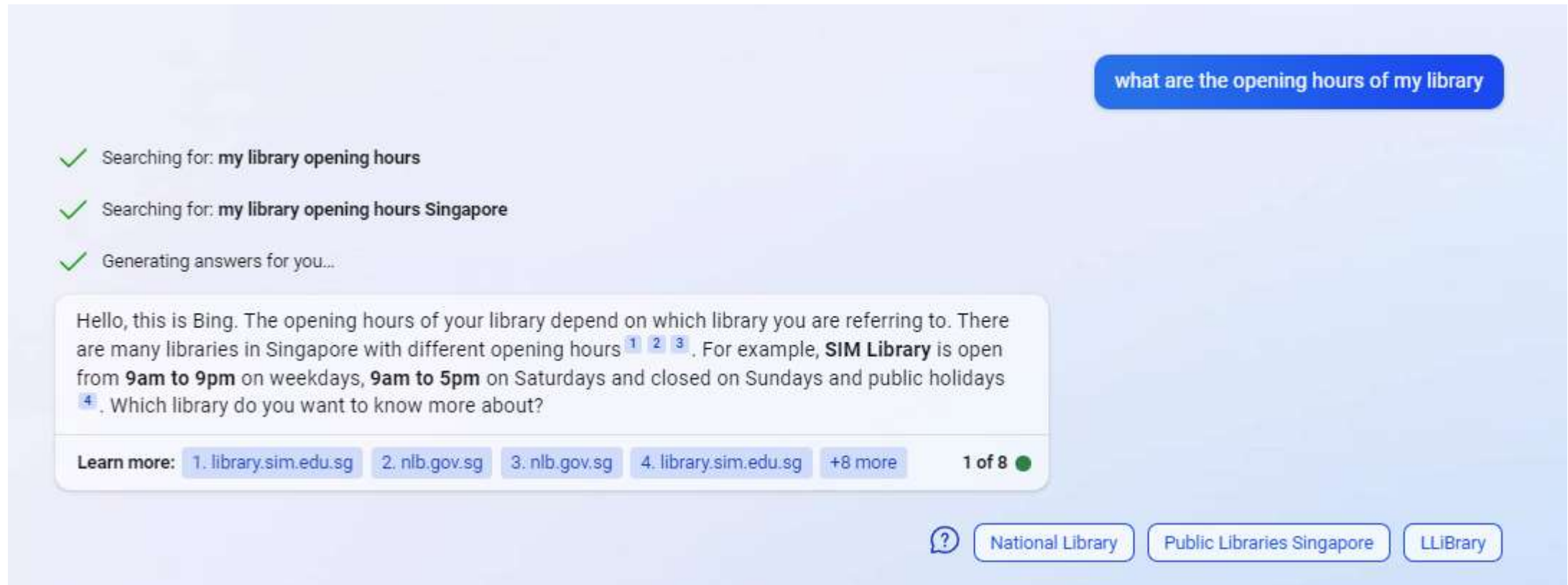
4. (PDF) About the size of Google Scholar: playing the numbers
PDF | On Jul 23, 2014, Enrique Orduña-Malea and others published About the size of Google Scholar: playing the numbers | Find, read and cite all the ...

1 science 2 springer 3 arxiv 4 researchgate

How Perplexity works (roughly)

- Step 1 – Finds relevant documents
- Step 2 – Extracts the most likely relevant passages
- Step 3 – Sends the extracted passages + prompt to language model (e.g. ChatGPT) with the instruction to answer the query using extracted passages

Bing+chat is a decent chatbot



The screenshot displays the Bing+chat interface. At the top right, a blue button contains the text "what are the opening hours of my library". Below this, three green checkmarks indicate the chatbot's processing steps: "Searching for: my library opening hours", "Searching for: my library opening hours Singapore", and "Generating answers for you...". The main chat area shows a response from Bing: "Hello, this is Bing. The opening hours of your library depend on which library you are referring to. There are many libraries in Singapore with different opening hours ¹ ² ³. For example, **SIM Library** is open from **9am to 9pm** on weekdays, **9am to 5pm** on Saturdays and closed on Sundays and public holidays ⁴. Which library do you want to know more about?". Below the text is a "Learn more:" section with four links: "1. library.sim.edu.sg", "2. nlb.gov.sg", "3. nlb.gov.sg", and "4. library.sim.edu.sg", followed by "+8 more" and "1 of 8" with a green dot. At the bottom right, there is a question mark icon and three buttons: "National Library", "Public Libraries Singapore", and "LLiBrary".

what are the opening hours of my library

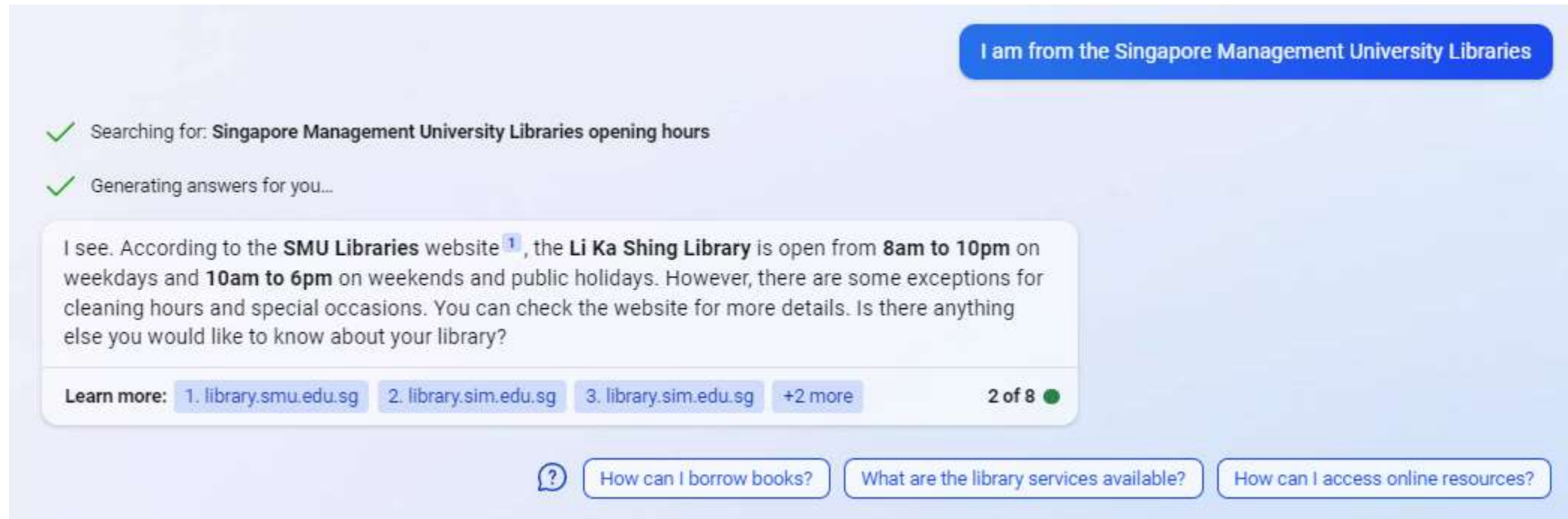
- ✓ Searching for: **my library opening hours**
- ✓ Searching for: **my library opening hours Singapore**
- ✓ Generating answers for you...

Hello, this is Bing. The opening hours of your library depend on which library you are referring to. There are many libraries in Singapore with different opening hours ¹ ² ³. For example, **SIM Library** is open from **9am to 9pm** on weekdays, **9am to 5pm** on Saturdays and closed on Sundays and public holidays ⁴. Which library do you want to know more about?

Learn more: [1. library.sim.edu.sg](#) [2. nlb.gov.sg](#) [3. nlb.gov.sg](#) [4. library.sim.edu.sg](#) +8 more 1 of 8 ●

? National Library Public Libraries Singapore LLiBrary

Bing+chat is a decent chatbot



I am from the Singapore Management University Libraries

✓ Searching for: **Singapore Management University Libraries opening hours**

✓ Generating answers for you...

I see. According to the **SMU Libraries** website ¹, the **Li Ka Shing Library** is open from **8am to 10pm** on weekdays and **10am to 6pm** on weekends and public holidays. However, there are some exceptions for cleaning hours and special occasions. You can check the website for more details. Is there anything else you would like to know about your library?

Learn more: [1. library.smu.edu.sg](#) [2. library.sim.edu.sg](#) [3. library.sim.edu.sg](#) +2 more 2 of 8

? How can I borrow books? What are the library services available? How can I access online resources?

Bing is a decent chatbot

I am a undergraduate how many books can I borrow?

✓ Searching for: **SMU Libraries borrowing books**

✓ Generating answers for you...

As an undergraduate student, you can borrow up to **20 books** at a time from the SMU Libraries ¹. The loan period is **28 days** for most books, but it may be shortened if another user requests the item ¹. You can borrow library materials using the self-checkout stations on Level 2 and 4 of the Li Ka Shing Library ². You will need your SMU card to scan the barcode of the items you want to borrow ². Do you have any other questions about borrowing books?

Learn more: [1. library.smu.edu.sg](https://library.smu.edu.sg) [2. libfaq.smu.edu.sg](https://libfaq.smu.edu.sg) [3. library.smu.edu.sg](https://library.smu.edu.sg)

3 of 8 ●



How can I renew books?

How can I request books?

What are the fines for overdue books?

Importance of getting your info indexed

Do I have access to the book "The book of why"?

- ✓ Searching for: **SMU Libraries The book of why**
- ✓ Searching for: **SMU Libraries The book of why Judea Pearl**
- ✓ Generating answers for you...

The book of why by Judea Pearl and Dana Mackenzie is a popular book about causal thinking and how it revolutionized science and the world ¹. Unfortunately, it seems that SMU Libraries do not have a physical or electronic copy of this book in their collections ² ³ ⁴. You may want to try other libraries or online sources to access this book. Alternatively, you can suggest SMU Libraries to purchase this book through their website ².

Learn more: [1. penguin.co.uk](#) [2. library.smu.edu.sg](#) [3. library.smu.edu.sg](#) [+3 more](#) 4 of 8 ●

[? Yes, please help me suggest SMU Libraries to purchase this book.](#) [No, thank you. I will look for other sources.](#)

Importance of getting your info indexed

How about the national library board of singapore?

- ✓ Searching for: **National Library Board Singapore The book of why**
- ✓ Generating answers for you...

The National Library Board (NLB) of Singapore is a network of libraries across Singapore that promotes reading, learning and history ¹. The NLB has an electronic copy of The book of why by Judea Pearl and Dana Mackenzie available for borrowing through their OverDrive service ². You will need a valid myLibrary ID to access this service. You can also search for other books and resources using their OneSearch platform ³.

Learn more: [1. nlb.gov.sg](https://www.nlb.gov.sg) [2. nlb.overdrive.com](https://www.nlb.gov.sg/overdrive) [3. search.nlb.gov.sg](https://www.nlb.gov.sg/one-search) +2 more

5 of 8 ●

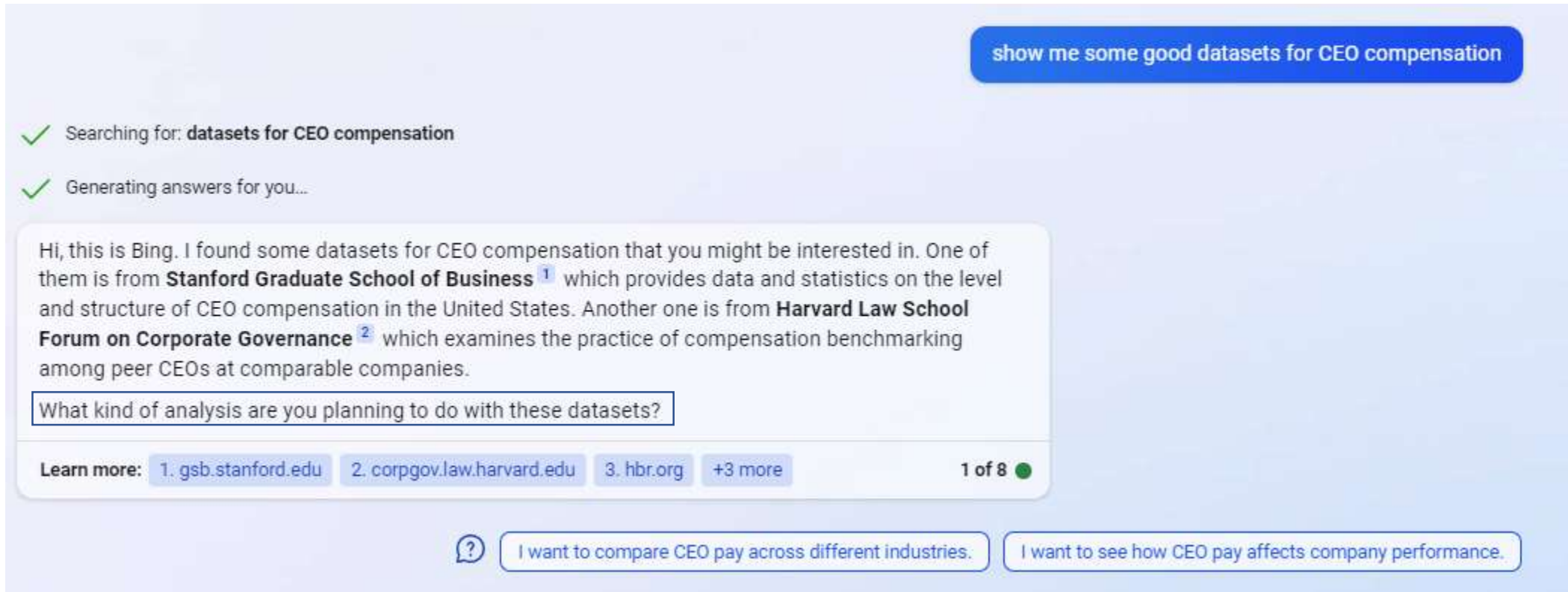


Yes, I have a myLibrary ID.

No, I don't have a myLibrary ID.

How can I get a myLibrary ID?

It can answer research questions



show me some good datasets for CEO compensation

✓ Searching for: **datasets for CEO compensation**

✓ Generating answers for you...

Hi, this is Bing. I found some datasets for CEO compensation that you might be interested in. One of them is from **Stanford Graduate School of Business** ¹ which provides data and statistics on the level and structure of CEO compensation in the United States. Another one is from **Harvard Law School Forum on Corporate Governance** ² which examines the practice of compensation benchmarking among peer CEOs at comparable companies.

What kind of analysis are you planning to do with these datasets?

Learn more: [1. gsb.stanford.edu](https://gsb.stanford.edu) [2. corpgov.law.harvard.edu](https://corpgov.law.harvard.edu) [3. hbr.org](https://hbr.org) +3 more 1 of 8 ●

? I want to compare CEO pay across different industries. I want to see how CEO pay affects company performance.

Follow ups

I want to see how CEO pay affects company performance.

✓ Searching for: **CEO pay and company performance**

✓ Generating answers for you...

That's an interesting question. There are different ways to measure CEO pay and company performance, and different studies may have different findings. One way to measure CEO pay is by using **realizable pay**, which is the value of compensation that CEOs actually receive based on stock price changes ¹. One way to measure company performance is by using **total shareholder return (TSR)**, which is the change in stock price plus dividends over a period of time ¹.

According to a study by **Pay Governance**, there is a strong correlation between CEO realizable pay and company TSR, meaning that CEOs who deliver higher returns for shareholders tend to receive higher compensation ¹. However, this may not be true for all industries or time periods.

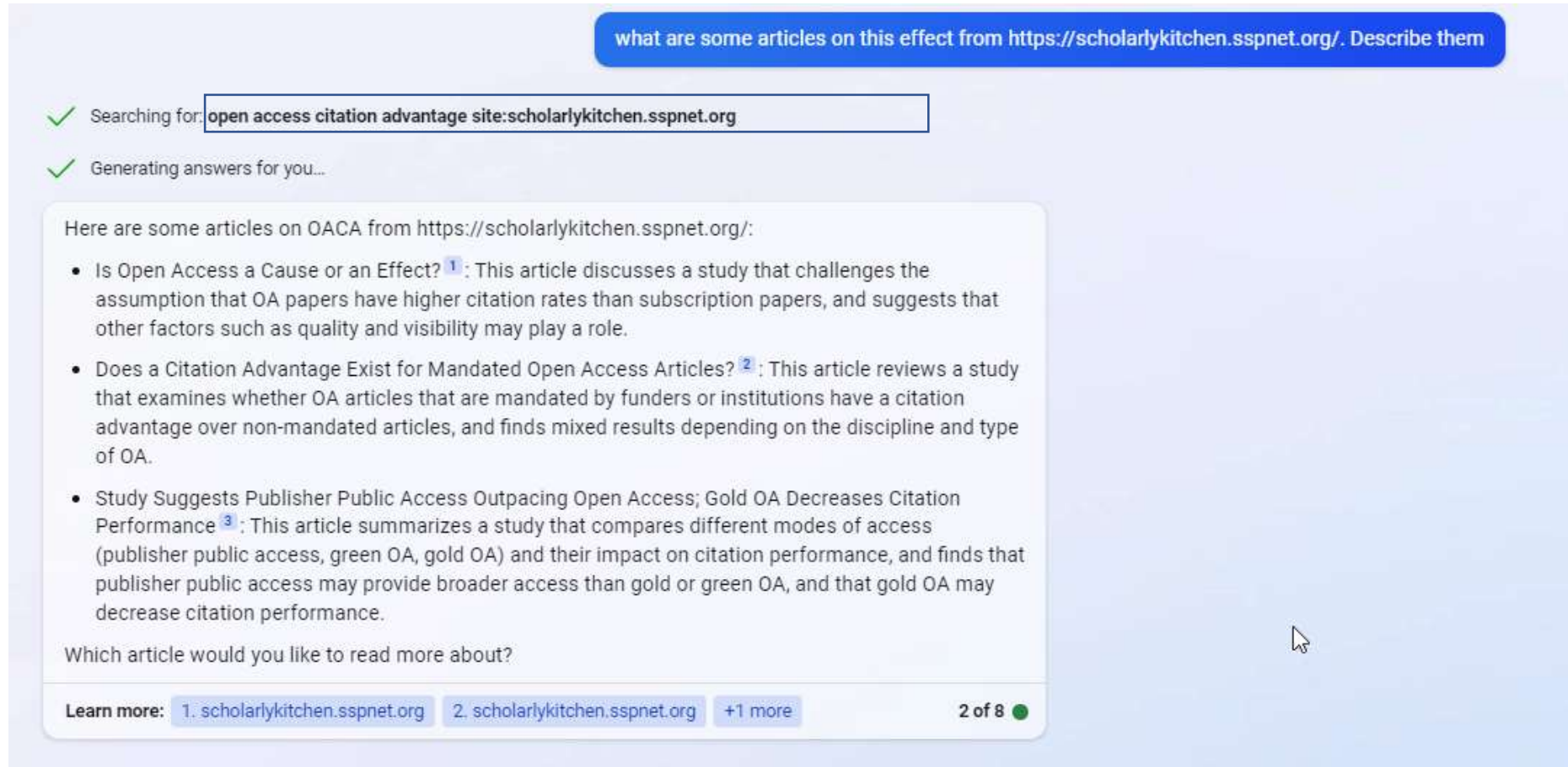
What industry or time period are you interested in?

Learn more: [1. corpgov.law.harvard.edu](#) [2. investopedia.com](#) [3. bdo.com](#) [+2 more](#)

2 of 8 ●



I can ask results to be from a given domain



what are some articles on this effect from <https://scholarlykitchen.sspnet.org/>. Describe them

✓ Searching for

✓ Generating answers for you...

Here are some articles on OACA from <https://scholarlykitchen.sspnet.org/>:

- Is Open Access a Cause or an Effect? ¹: This article discusses a study that challenges the assumption that OA papers have higher citation rates than subscription papers, and suggests that other factors such as quality and visibility may play a role.
- Does a Citation Advantage Exist for Mandated Open Access Articles? ²: This article reviews a study that examines whether OA articles that are mandated by funders or institutions have a citation advantage over non-mandated articles, and finds mixed results depending on the discipline and type of OA.
- Study Suggests Publisher Public Access Outpacing Open Access; Gold OA Decreases Citation Performance ³: This article summarizes a study that compares different modes of access (publisher public access, green OA, gold OA) and their impact on citation performance, and finds that publisher public access may provide broader access than gold or green OA, and that gold OA may decrease citation performance.

Which article would you like to read more about?

Learn more: [1. scholarlykitchen.sspnet.org](https://scholarlykitchen.sspnet.org/) [2. scholarlykitchen.sspnet.org](https://scholarlykitchen.sspnet.org/) [+1 more](#) 2 of 8 ●

Restricting to OA papers!

The world's largest collection of research papers

Search 259M from

SEARCH

Bronze, Free or Fourrée?

Following a study of over 300,000 articles, Piwowar et al. [8] posited the term "bronze":

We also add a novel subcategory, Bronze. Bronze shares attributes of Gold and Hybrid; like both, Bronze OA articles are publisher-hosted. Unlike Gold OA, Bronze articles are not published in journals considered open access in the DOAJ. Unlike Hybrid, Bronze articles carry no license information. Although this lack of identifiable license may not be intentional, without an identifiable license, the articles are free to read but do not allow extended reuse rights beyond reading.

Notably, the term "bronze access" also appears in a tweet from Ridgway as far back as 2014 [9].

A central finding of the Piwowar study was their suggestion of strong evidence for the existence of an Open Access Citation Advantage. Open Access Citation Advantage postulates that open access articles get cited more than their closed counterparts. It is difficult to prove this decisively in the absence of randomized controlled trials, as authors could conceivably only choose to publish their best work open access, particularly if high author fees are involved, but increasing evidence points in this direction [10]. Moreover, all other things being equal, it is hard to see how articles that have limited access can be read and cited as widely as those with no access restrictions.

So-called bronze articles also enjoy the benefit of greater citation. Piower et al. [8] found that half of bronze articles ap-

providing access to the open access citation advantage that has not been hitherto considered by researchers; i.e., that technologies generally found alongside Creative Commons may aid article indexation and discoverability. However, this is not to imply that some parts of the world are necessarily playing catch-up to others regarding open access. Many journals in the global north remain fully closed or even offline. Meanwhile, initiatives such as the SciELO (Scientific Electronic Library On-line) project in Brazil have made massive advances in open access publishing in Latin America and the Caribbean before spreading to Portugal, Spain, and South Africa [12].

However, another of Piwowar et al. [8]'s significant findings was a high prevalence of bronze articles in many traditional closed and hybrid journals. This would appear to be an increasing trend with major publishers, but not a lot is known about this phenomenon. A casual perusal of the prominent hybrid journals from the biggest commercial publishers can reveal that several articles are marked as "free." Yet this "free" label may mean that it is free to read only on the journal website. The publisher can, in theory, revoke this access at any time. Harnad [13] has derided this practice as "peek-a-boo open access." The value of this free access is clear to the publisher: It can make some articles free, such as those that it thinks will drive readers, attention, and citations to its journal. This increases the reputation of the journal, but because it does not make all of its articles free all of the time, it can still charge subscription fees to institutions and readers.

These "free" or Bronze open access articles grant no reuse rights. It may not even be clear that the articles can be legally downloaded from the journal website and retained. There are certainly no rights to share or redistribute them. Combined



✓ Searching

✓ Generating

The term "bronze access" is only from core.ac.uk

Learn more:



What is Gold OA?

What is Hybrid OA?

What is Green OA?

Compare and contrast methods!

find for me papers that estimate the size of google scholar and compare and contrast the approaches

- ✓ Searching for: papers that estimate the size of google scholar
- ✓ Searching for: compare and contrast the approaches to estimate the size of google scholar
- ✓ Generating answers for you..

There are several papers that estimate the size of Google Scholar using different approaches. One paper¹ proposes four methods: direct estimation, indirect estimation, extrapolation and comparison. Another paper² compares 12 academic databases using a method based on stratified sampling and capture-recapture analysis. The main differences between these approaches are:

- Direct estimation involves querying Google Scholar with specific terms and counting the number of results returned.
- Indirect estimation involves querying Google Scholar with terms that are known to appear in a certain number of documents and calculating the ratio of results returned to documents expected.
- Extrapolation involves querying Google Scholar with terms that have a known frequency distribution in a large corpus and estimating the size based on the observed frequencies.
- Comparison involves comparing the coverage of Google Scholar with other databases that have a known size or an estimate of their size.
- Stratified sampling involves dividing the academic web into strata based on disciplines, languages and document types and randomly sampling documents from each stratum.
- Capture-recapture analysis involves estimating the size based on the overlap between two independent samples.

What aspect of these approaches are you most interested in?

Learn more: [1. arxiv.org](#) [2. link.springer.com](#) [3. scholar.google.com](#) [+2 more](#) 1 of 10 ●

The accuracy and reliability of each approach. The advantages and disadvantages of each approach. The applicability and limitations of each approach.

Academic Search enhanced by Large Language Models

The logo for Elicit, featuring the word "Elicit" in a light blue, sans-serif font.

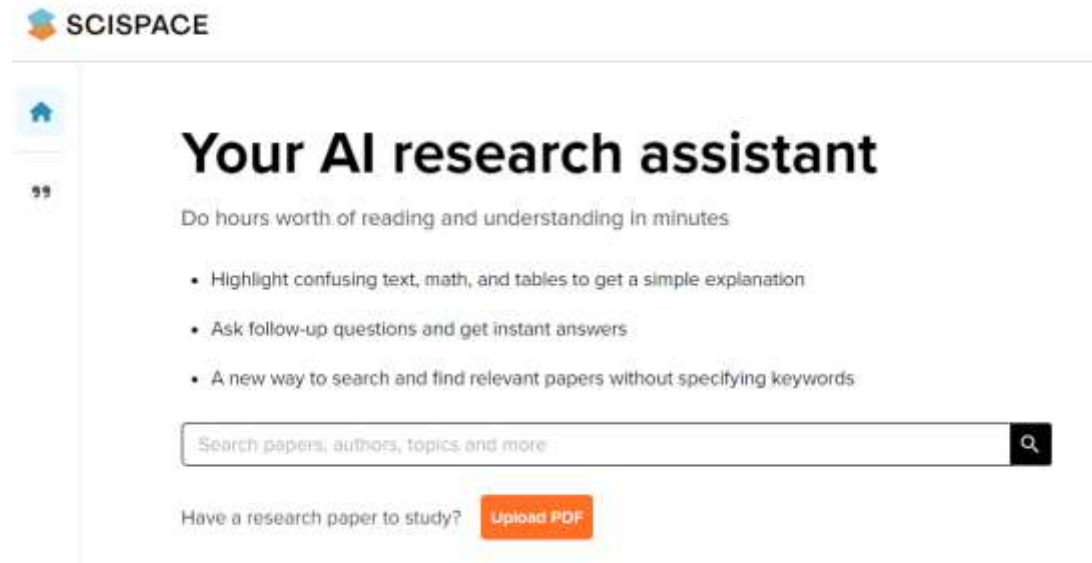
<https://elicit.org/>

The logo for Scite, featuring the word "scite_" in white, lowercase, sans-serif font on a blue rectangular background.

<https://scite.ai/>

The logo for Consensus, featuring a stylized "C" made of two overlapping shapes (blue and green) and the word "consensus" in a black, sans-serif font.

<https://consensus.app/search/>

A screenshot of the SCISPACE website interface. The header includes the SCISPACE logo and a navigation menu with a home icon and a quote icon. The main heading is "Your AI research assistant" with the subtext "Do hours worth of reading and understanding in minutes". Below this is a list of features: "Highlight confusing text, math, and tables to get a simple explanation", "Ask follow-up questions and get instant answers", and "A new way to search and find relevant papers without specifying keywords". There is a search bar with the placeholder text "Search papers, authors, topics and more" and a search icon. At the bottom, there is a link "Have a research paper to study?" and an orange "Upload PDF" button.

<https://typeset.io/>

<https://musingsaboutlibrarianship.blogspot.com/2022/11/q-academic-systems-elicitorg-scispace.html>

Scite- beta –Answer a question

The screenshot displays the Scite website interface. At the top left is the 'scite_' logo. A search bar contains the query 'can i use google scholar alone for systematic reviews?'. To the right of the search bar are links for 'Install extension!' and 'Product'. Below the search bar, there are tabs for 'Search', 'All', 'Citation Statements', 'Papers', 'Ask a Question', and 'Beta'. A 'Share Results' button is located to the right of the tabs. Below the tabs, there are dropdown menus for 'Year' and 'Citations', and links for 'What is Ask a Question' and 'Provide feedback'. The main content area features an 'Answer Summary' section with an information icon. The summary text reads: 'No, it is not recommended to use Google Scholar alone for systematic reviews. While one study has shown that Google Scholar alone can retrieve all necessary sources for top systematic reviews in medicine [2], other sources indicate that Google Scholar should not be used in isolation, especially when searching for patient care information or literature reviews [4]. However, Google Scholar can be used in addition to other reliable sources of information [3][5]. It is important to note that systematic reviews should be given the status and time allocation of a significant research project [1].'

scite_ can i use google scholar alone for systematic reviews? Install extension! Product Re

Search All Citation Statements Papers Ask a Question Beta Share Results

Year Citations What is Ask a Question Provide feedback

Answer Summary ⓘ

No, it is not recommended to use Google Scholar alone for systematic reviews. While one study has shown that Google Scholar alone can retrieve all necessary sources for top systematic reviews in medicine [2], other sources indicate that Google Scholar should not be used in isolation, especially when searching for patient care information or literature reviews [4]. However, Google Scholar can be used in addition to other reliable sources of information [3][5]. It is important to note that systematic reviews should be given the status and time allocation of a significant research project [1].

Scite- beta –Answer a question




Reference # 2

“...The comprehensiveness and easy accessibility of Google searches can be used to detect plagiarised sentences and larger portions of text, particularly in the absence of a specialised plagiarism-detecting software (15). A study comparing Google Scholar with PubMed and Cochrane Library searches for coverage of the literature for top systematic reviews in medicine proved that searches through Google Scholar alone are sufficient for retrieving all the necessary sources (16). ...”

Section: Google Scholar

Multidisciplinary Bibliographic Databases

[Armen Yuri Gasparyan](#)¹, [Lilit Ayvazyan](#)², [George D. Kitas](#)³ 2013 *J Korean Med Sci*

 48 |  0 |  43 |  0

[View full text](#) [Add to dashboard](#) [Buy / Rent full text](#)

Elicit – answers the question

- <https://elicit.org/search?q=can+you+use+google+scholar+alone+for+systematic+reviews%3F>

The screenshot displays the Elicit search interface. At the top, the search bar contains the query "can you use google scholar alone for systematic reviews?". Below the search bar, the results are organized into two main sections. The left section, titled "can you use google scholar alone for systematic reviews?", features a "SUMMARY OF TOP 4 PAPERS" section with a "BETA" label. The summary text reads: "These papers suggest that you cannot use Google Scholar alone for systematic reviews. Gehanno 2013 found that Google Scholar is sensitive enough to be used alone for systematic reviews, but Giustini 2013 found that it is not enough to be used alone. Gusenbauer 2020 found that only a few search engines can be recommended for evidence syntheses without adding substantial caveats, and Falagas 2007 found that PubMed remains an optimal tool in biomedical electronic research. Therefore, it is not advisable to use Google". The right section, titled "Paper title", lists two papers. The first paper is "Is the coverage of google scholar sufficient for systematic reviews?" by Jean-François Gelinas, published in BMC Medical Information Research and Open Access in 2013, with 220 citations. The second paper is "Google Scholar is not recommended for evidence syntheses without adding substantial caveats" by D. Giustini and M. Bouillon-Buonafina, published in the Online Journal of Critical Appraisal in 2013, as a Systematic Review. Each paper entry includes a star icon for favoriting and a trash icon for deletion.

Elicit extracts characteristics of papers

Elicit [FAQ](#) [Tasks](#) [Starred](#) [Settings](#)

can you use google scholar alone for systematic reviews?

SUMMARY OF TOP 4 PAPERS BETA

These papers suggest that you cannot use Google Scholar alone for systematic reviews. [Gehanno 2013](#) found that Google Scholar is sensitive enough to be used alone for systematic reviews, but [Giustini 2013](#) found that it is not enough to be used alone. [Gusenbauer 2020](#) found that only a few search engines can be recommended for evidence syntheses without adding substantial caveats, and [Falagas 2007](#) found that PubMed remains an optimal tool in biomedical electronic research. Therefore, it is not advisable to use Google

Add information about all papers

- Outcomes measured
- Number of participants
- Main findings**

Search for paper information

Paper title	Main findings	Detailed study design
<p>★ Is the coverage of google scholar enough to be used alone for systematic reviews</p> <p>✉ Jean-François Gehanno, Laetitia Rollin, Stéfan Jacques Darmoni</p> <p>📄 <i>BMC Medical Informatics and Decision Making</i></p> <p>2013 Review 220 Citations DOI PDF</p>	<ul style="list-style-type: none"> • All the 738 original studies included in the gold standard database were retrieved in GS (100%). • GS could become the leading bibliographic database in medicine and could be used alone for systematic reviews. • With some improvement in the research options, to increase its precision, GS could become the leading bibliographic database in medicine and could be used alone for systematic reviews. 	<p>a retrospective analysis of existing systematic reviews</p>
<p>★ Google Scholar is not enough to be used alone for systematic reviews</p> <p>✉ D. Giustini, M. Boulos</p> <p>📄 <i>Online Journal of Public Health Informatics</i></p> <p>2013 Systematic Review 131 Citations DOI PDF</p>	<ul style="list-style-type: none"> • Google Scholar produced records for ~95% of the papers cited by Chou et al. (n=476/506). • Google Scholar is not suitable for systematic reviews due to its constantly-changing content, algorithms and database structure. • Google should provide more details about its database coverage and improve its interface. 	<p>a comparison of Google Scholar and other academic search engines</p>
<p>★ Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources</p> <p>✉ Michael Gusenbauer, Neal R Haddaway</p> <p>📄 <i>Research Synthesis Methods</i></p> <p>2020 Review 435 Citations DOI PDF</p>	<ul style="list-style-type: none"> • There are substantial differences in the performance of search systems, meaning that their usability in systematic searches varies. • Only half of the search systems analyzed and only a few Open Access databases can be recommended for evidence syntheses without adding substantial caveats. • Google Scholar is inappropriate as a principal search system for evidence synthesis. 	<p>a novel, query based method that tested how well users were able to interact and retrieve records with each system</p>

You can use the preset columns or add your own

Number of studies

Population characteristics

Population summary

Age

Organism

Region

INTERVENTION STUDIED

Intervention

Dose

Duration

Q What was the...

Paper title	Abstract summary	What was the dataset used?
<p>★ The citation advantage of open-access articles</p> <p>Michael Norris, C. Oppenheim, Fytton Rowland</p> <p><i>J. Assoc. Inf. Sci. Technol.</i></p> <p>2008 502 Citations DOI PDF</p>	<p>There appears to be a clear citation advantage for those articles that are OA as opposed to those that are TA.</p>	<p>OAister, OpenDOAR, and RePEC</p>
<p>★ Does open access citation advantage depend on paper topics?</p> <p>H. Sotudeh</p> <p><i>Journal of information science</i></p> <p>2020 12 Citations DOI</p>	<p>The OA citation surplus is not an artefact of the OA and NOA papers' topics and, therefore, their citation potential.</p>	<p>a sample of papers published in 47 Elsevier article processing charges (APC) funded journals in different access models including non-open access (NOA), APC, Green and mixed Green APC</p>
<p>★ The citation advantage of open-access articles</p> <p>Michael Norris, Charles Oppenheim, Fytton Rowland</p> <p>2008 124 Citations DOI</p>	<p>There appears to be a clear citation advantage for those articles that are OA as opposed to those that are TA.</p>	<p>a sample of 4,633 articles</p>
<p>★ Analysis on open access citation advantage: an empirical study based on Oxford open journals</p> <p>Lifang Xu, Jinhong Liu, Qing Fang</p> <p><i>JConference</i></p> <p>2011 19 Citations DOI</p>	<p>There is citation advantage for open access articles published in Oxford Open journals over the non-OA ones.</p>	<p>12,354 original research articles published in 93 Oxford Open journals in 2009</p>

You can check the passage it uses to answer the question

D. Giustini, M. Boulos
 Online Journal of Public Health Informatics
 2013 131 Citations PDF Semantic Scholar DOI

Abstract summary
 Google Scholar's constantly-changing content, algorithms, and database structure make it a poor choice for systematic reviews.

Who were the participants?

- France
- France

Main findings

- Google Scholar produced records for ~95% of the papers cited by Chou et al. (n=476/506).
- Google Scholar is not suitable for systematic reviews due to its constantly-changing content, algorithms and database structure.
- Google should provide more details about its database coverage and improve its interface.

Can I trust this paper?

- This study was a comparison of Google Scholar and other academic search engines.

Ask a question about this paper **Ask**

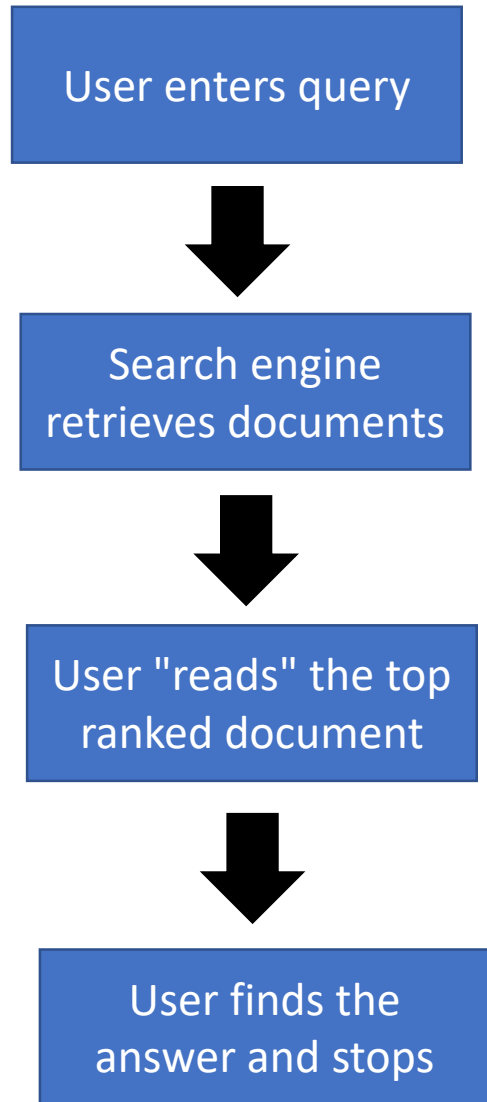
In 2013, French researchers, Gehanno et al., published a study that asked a simple question to which most expert searchers thought they knew the answer: 'Is Google scholar enough to be used alone for systematic reviews?' [16] The authors state that GS coverage has improved and ask whether its "coverage is high enough to be used alone in systematic reviews". In other words, the authors ask whether GS might replace MEDLINE and other bibliographic databases to perform costly, time-intensive searches for systematic reviews. The clearly-stated question and OJPHI conclusions of Gehanno et al. are examined in this paper; we ask whether Google Scholar has improved enough over the years to be used alone in systematic reviews.

Methods

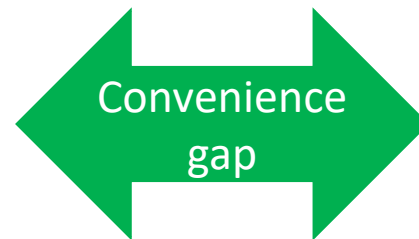
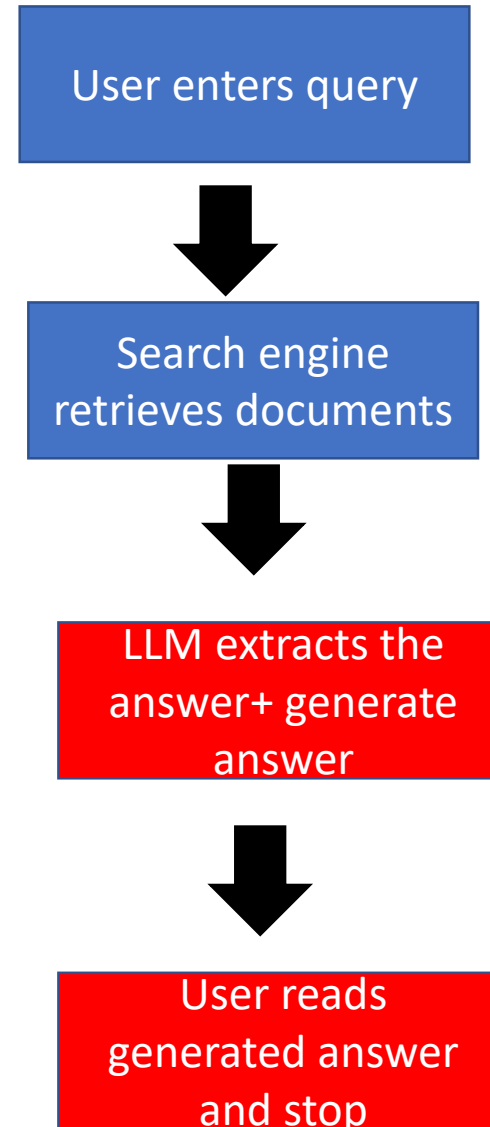
The authors searched for a systematic review that was comparable in size to Gehanno et al. We selected a recent study in our area of expertise (health/public health informatics), Chou et al. (2013), and contacted the authors for a list of the 506 publications they found in their SR on social media in health 1. To test Google Scholar's ability to locate articles from an existing systematic review, we searched for all of the publications found by Chou et al. [17].

We tested whether the 500+ articles that formed the basis of Chou et al.'s SR were indexed by GS. Since we knew what we were looking for, and were not testing GS' ability to produce relevant documents, our searches were straightforward title searches. Chou et al. provided us with an Excel spreadsheet of the titles of papers (n=514) that comprised their systematic review. After correcting for minor errors, we looked for 506 unique items occurring either as simple citations or full-text links to papers within GS. We checked for the presence of these 506 publications by querying GS for the title of each study (in quotes ""), one by one. When a search failed to retrieve the required article, or produced too many results to browse, we opted to use Google's allintitle: command to increase our precision and search accuracy by limiting our search to the titles of articles. Some papers that were not found in GS were later searched and found in regular Google Search. Our

Before



After



Some questions & speculations

Will people bother to verify citations?

Ranking of results in search engines are even more important?

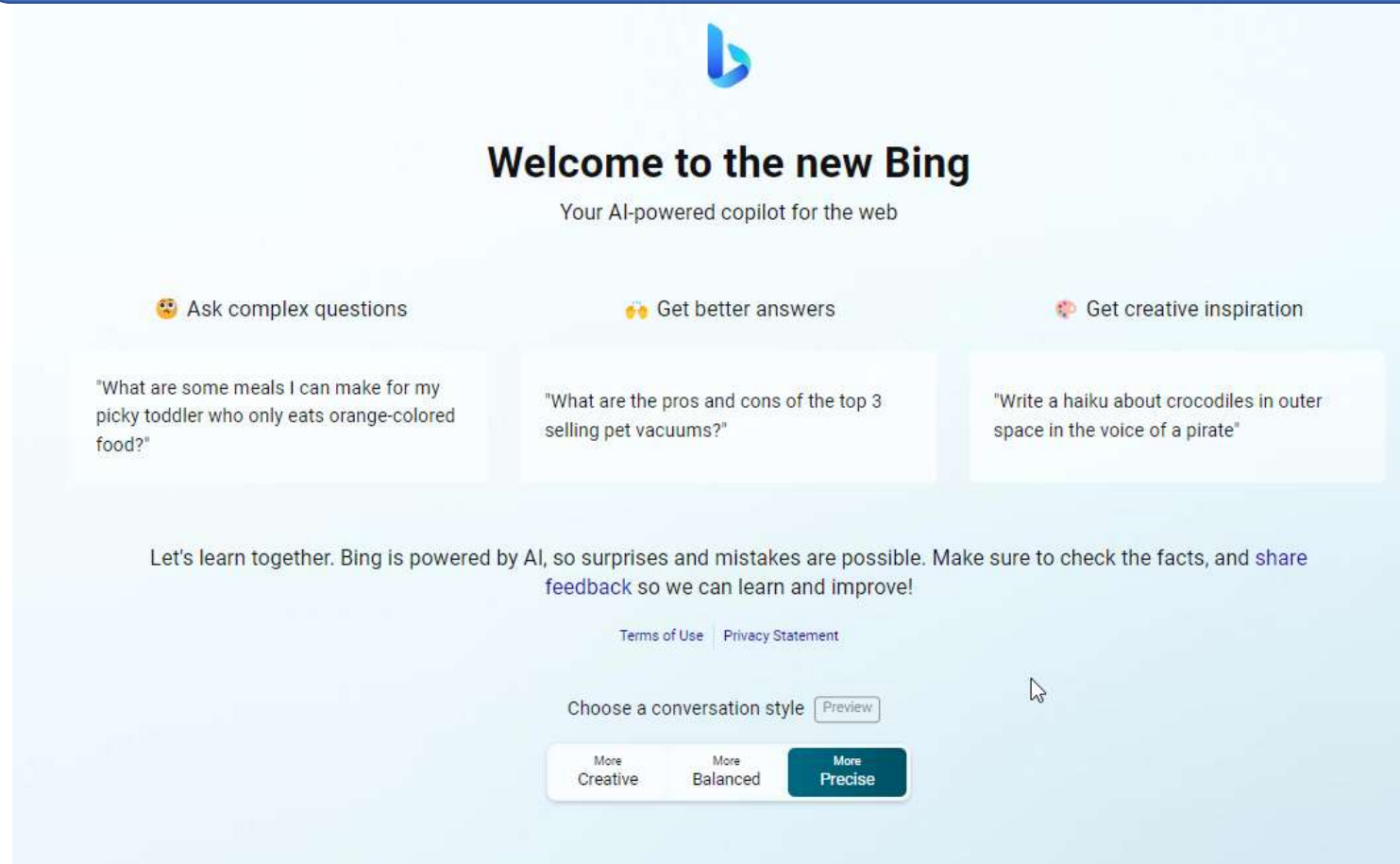
Open or Die?

Will over-reliance on summaries lead to superficial learning

Importance of deep expertise

Some questions & speculations

Will people bother to verify citations?



The screenshot shows the new Bing homepage. At the top center is the Bing logo, a stylized blue 'b'. Below it, the text reads "Welcome to the new Bing" in a bold, black font, followed by "Your AI-powered copilot for the web" in a smaller, regular font. There are three main feature cards arranged horizontally. The first card is titled "Ask complex questions" with a question mark icon and contains the text: "What are some meals I can make for my picky toddler who only eats orange-colored food?". The second card is titled "Get better answers" with a lightbulb icon and contains the text: "What are the pros and cons of the top 3 selling pet vacuums?". The third card is titled "Get creative inspiration" with a paint palette icon and contains the text: "Write a haiku about crocodiles in outer space in the voice of a pirate". Below these cards, a message states: "Let's learn together. Bing is powered by AI, so surprises and mistakes are possible. Make sure to check the facts, and share feedback so we can learn and improve!". At the bottom, there are links for "Terms of Use" and "Privacy Statement". A section titled "Choose a conversation style" includes a "Preview" button and three options: "More Creative", "More Balanced", and "More Precise", with the "More Precise" option highlighted in a dark teal color.

Some questions & speculations

Ranking of results in search engines are even more important?

"Our approach also raises a challenging problem with societal implications: how should factual accuracy be evaluated when training AI systems? Evans et al. [2021, Section 2] propose a number of desiderata, but a substantial gap remains between these and the highly specific criteria needed to train current AI systems with reasonable data efficiency. **We made a number of difficult judgment calls, such as how to rate the trustworthiness of sources** (see Appendix C), which we do not expect universal agreement with. While WebGPT did not seem to take on much of this nuance, we expect these decisions to become increasingly important as AI systems improve, and think that **cross-disciplinary research is needed to develop criteria that are both practical and epistemically sound**"

[WebGPT: Browser-assisted question-answering with human feedback](#)

Some questions & speculations

Ranking of results in search engines are even more important?



how do you find seminal papers?

✓ Searching for: **how to find seminal papers**

✓ Generating answers for you...

Seminal papers are papers that have made a significant impact on a field of study. One way to find them is to use citation analysis tools such as Scopus ¹ or Google Scholar ² that show how many times a paper has been cited by other sources. A high number of citations usually indicates that a paper is influential and seminal.

What field of study are you interested in finding seminal papers for?

Learn more: [1. mcphs.libguides.com](#) [2. subjectguides.lib.neu.edu](#) [+3 more](#) 1 of 8

Artificial intelligence. Psychology. History.

Ask me anything...
0/2000

How do you ask to see results from the next top ranked results?

Some questions & speculations

Open or Die?

- Open Access Citation advantage becomes stronger?
- Will publishers allow search engines like google scholar to index their full text behind paywalls?

what is the main finding of the first paper

The main finding of the first paper titled "Methods for estimating the size of Google Scholar" is that it proposes various methods to estimate the current size (number of indexed documents) of Google Scholar. These methods include an estimate based on empirical studies of Google Scholar coverage and two internal estimate methods based on direct, empty and absurd queries. Would you like more information on this topic?

4 of 10 ●



Yes, tell me more about these methods.

What are some other papers on this topic?

How accurate are these estimates?

Will publishers allow search engines to index their full-text behind paywall

Some questions & speculations

Will over-reliance on summaries lead to superficial learning

Importance of deep expertise

Further Implications

- User education

- Users should be aware of the weaknesses of generative AI
- Do such tools increase the efficiency and effectiveness of fake news?
- Long term effects on the use of such technology

- Reference work

- Will such technologies further reduce the need for reference desks?
- Are competent, easily maintained chatbots as first tier support now in sight?

- Others

- Should we value Open Access, OER, Open Science etc more?
- Are there roles we can play in supporting the development of open source LLMs? Supply of quality data?
- Training of such bots for assessment of creditability?

Should we double down on "Open"



Aaron Tay

Dec 6, 2022 · 14 min read · Listen



Are we undervaluing Open Access by not correctly factoring in the potentially huge impacts of Machine learning? — An academic librarian's view (I)

Synopsis : I have recently adjusted my view to the position that the benefits of Machine learning techniques are more likely to be real and large. This is based on the recent incredible results of LLM (Large Language models) and about a year's experimenting with some of the newly emerging tools based on such technologies.

If I am right about this, are we academic librarians systematically undervaluing Open Access by not taking this into account sufficiently when negotiating with publishers? Given that we control the purse strings, we are one of the most impactful parties (next to publishers and researchers) that will help decide how fast if at all the transition to an Open Access World occurs. For example, should we be willing to even pay a bit more so that we can transition to a Open Access world from a subscription world faster albeit at higher outflows from us? This can come in the form of the prices we pay in so called Transformative deals like Read and Publish deals or S2O etc.

<https://medium.com/a-academic-librarians-thoughts-on-open-access/are-we-undervaluing-open-access-by-not-correctly-evaluating-the-potentially-huge-impacts-of-e93af1de9414>

Questions for researchers – from Nature

Questions for debate

Issues for discussion at a forum about conversational AIs.

- Which research tasks should or should not be outsourced to large language models (LLMs)?
- Which academic skills and characteristics remain essential to researchers?
- What steps in an AI-assisted research process require human verification?
- How should research integrity and other policies be changed to address LLMs?
- How should LLMs be incorporated into the education and training of researchers?
- How can researchers and funders aid the development of independent open-source LLMs and ensure the models represent scientific knowledge accurately?
- What quality standards should be expected of LLMs (for example, transparency, accuracy, bias and source crediting) and which stakeholders are responsible for the standards as well as the LLMs?
- How can researchers ensure that LLMs promote equity in research, and avoid risks of widening inequities?
- How should LLMs be used to enhance principles of open science?
- What legal implications do LLMs have for scientific practice (for example, laws and regulations related to patents, copyright and ownership)?

<https://www.nature.com/articles/d41586-023-00288-7>

Questions for academic librarians

- Which library tasks should or should not be out-sourced to Large Language Models (LLMs)
- Which Library skills remain essential to Librarians?
- What steps in a AI-assisted library process require Librarian Verification?
- How can libraries aid the development of independent Open-Source? (LLMs)
- What legal implications do LLMs have for laws and regulations related to patents, copyright?
- How should librarians keep up?

A tentative reorganization of roles – Proposed 2018?

Collection of new inputs, processes, outputs –

Technical Service?

- Mgt of CRIS, Institutional & data repositories
- Managing OA / OER
- Curation of DH, DS projects
- Processing of open citations

Discovering & facilitating use

Reference?

- Discovery & use of open resources (e.g., OER, Wikidata, TDM sources)
- Processing/Manipulation (e.g. Digital Scholarship/Humanities e.g. TDM , GIS)
- Data carpentry

Education on workflow & new technologies

Information Literacy?

- Advice on the various “open”
 - Open Access
 - Open Data
 - Open Citations
 - Open Science
- **Fake news****
- **Impact of Algo****

Building Supporting Library & Scholarly Infrastructure – Analytics & ML (Library IT)

Let me end in the words of ChatGPT....

In conclusion, the rise of Generative AI, powered by technologies like large language models and text-to-image AI, is going to have a significant impact on academic libraries. But this impact is not something to fear; it's something to embrace. We have a unique opportunity to play a significant role in ensuring that AI algorithms have access to high-quality training data, which will help to improve their accuracy and effectiveness. As long as we are proactive in managing data and mindful of ethical considerations, academic libraries will continue to be at the forefront of research and discovery.

Thank you

Aaron Tay



Aarontay@gmail.com



[@aarontay](https://twitter.com/aarontay)



<https://musingsaboutlibrarianship.blogspot.com>