

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Library

SMU Libraries

9-2023

AI in Universities and Libraries

Aaron TAY

Follow this and additional works at: https://ink.library.smu.edu.sg/library_research



Part of the [Information Literacy Commons](#), and the [Scholarly Communication Commons](#)

This Conference Paper is brought to you for free and open access by the SMU Libraries at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Library by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylids@smu.edu.sg.

AI in Universities and Libraries –

EIFL2023

19 Sept 2023

Aaron Tay

Lead, Data Service

Singapore Management
University Libraries

AaronTay@smu.edu.sg

SMU Libraries organized a Hackathon on use of Large Language Models/Generative AI in Aug 2023



Welcome to SMU Libraries' inaugural hackathon Generate Your L(AI)brary, where technology enthusiasts and creative problem solvers come together to reimagine the future of libraries and research through the power of AI.

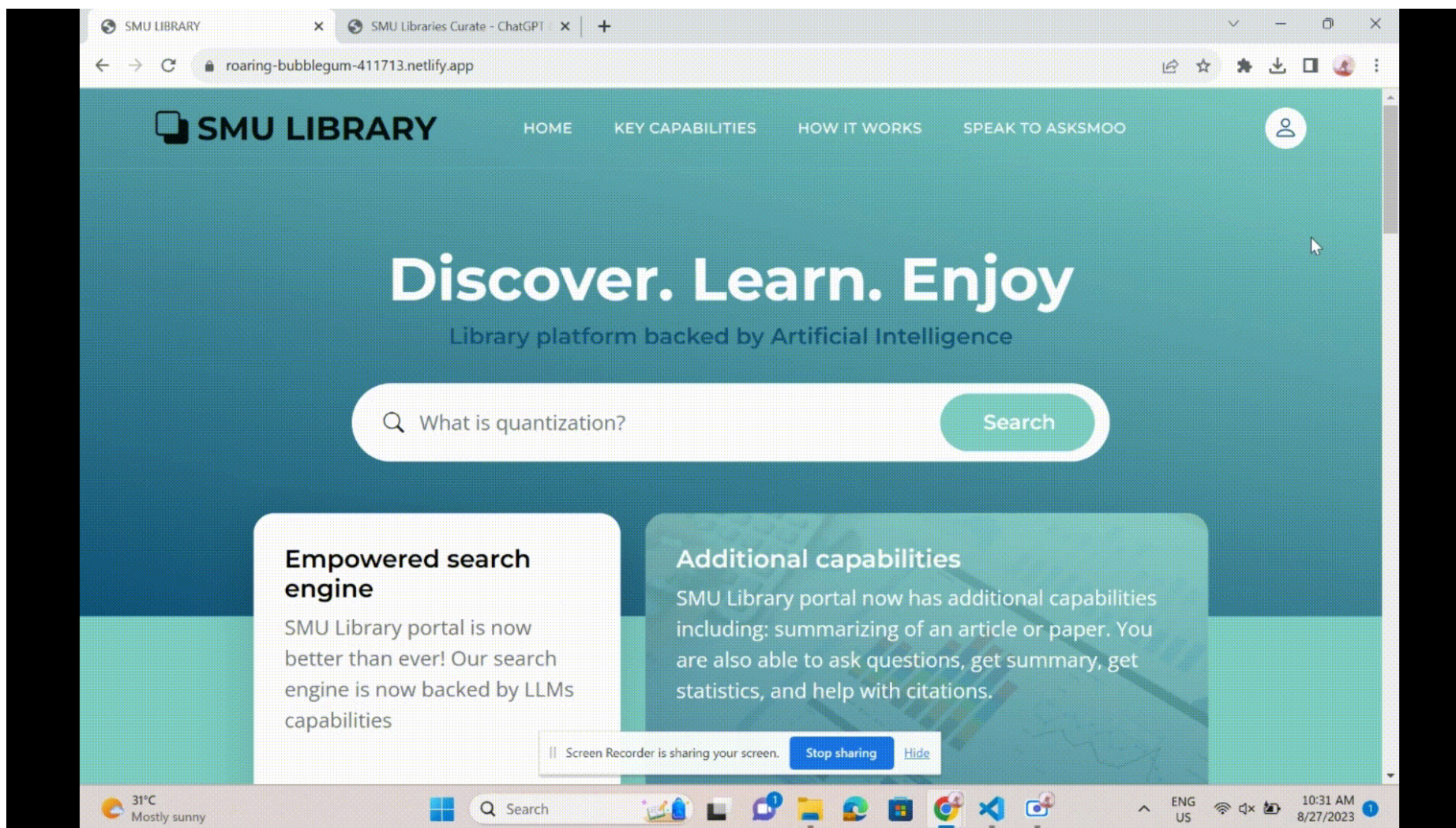
This year, our theme revolves around the concept of **"Refining library and research experience with GPT"**. To empower participants in their exploration and experimentation, **each participating team will receive sponsorship of up to USD 50 worth of OpenAI's API tokens (with access to GPT-4!)** for the entire duration of the hackathon.

Join us for an exhilarating 24 hours of collaboration and exploration to leverage the power of technology to craft creative applications that enhance your library and research experience!

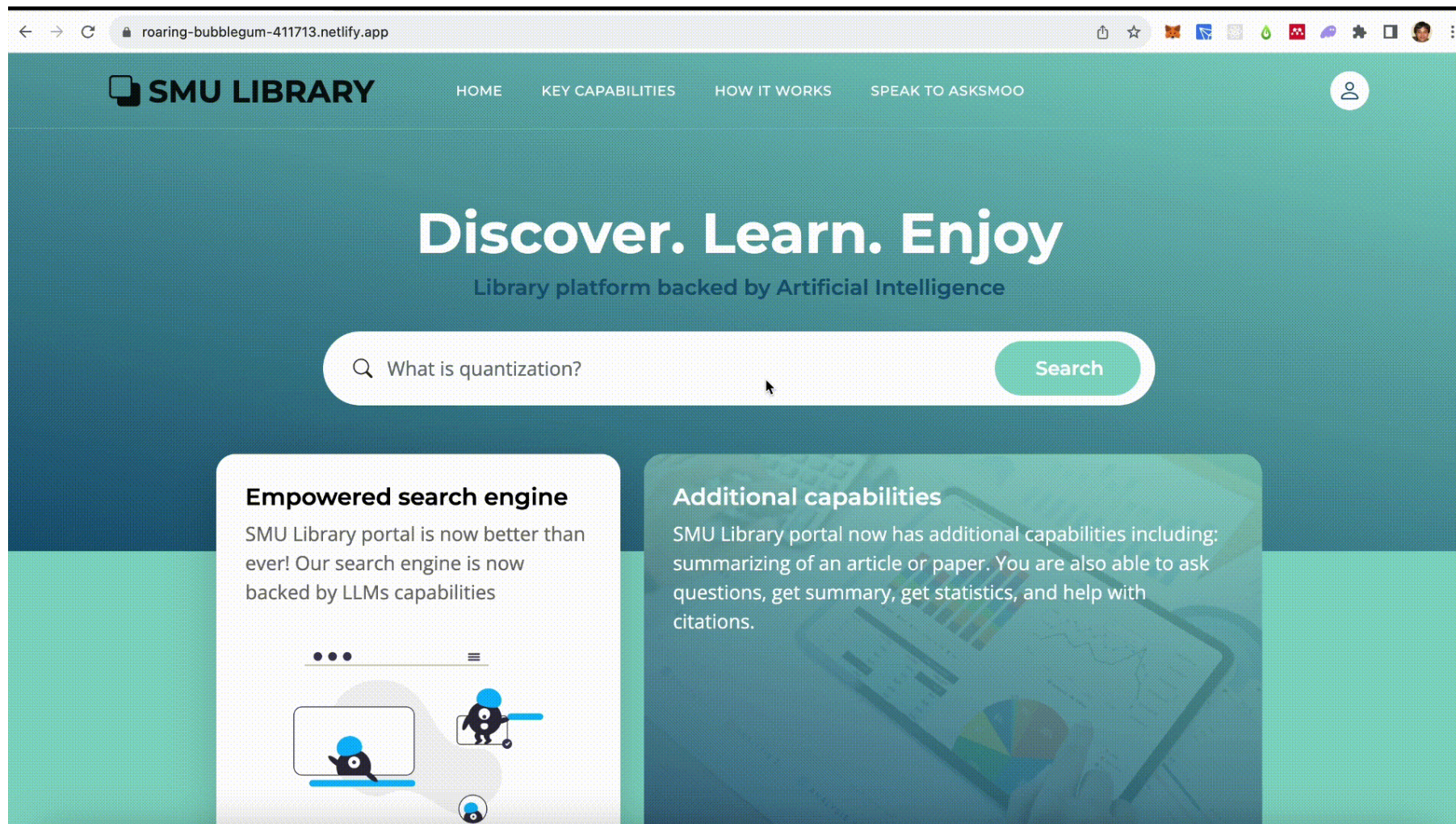
**What did the hackathon
participants come up
with?**



Search , and more Search (via Chatbot)



Search , Search and more Search (via Chatbot)

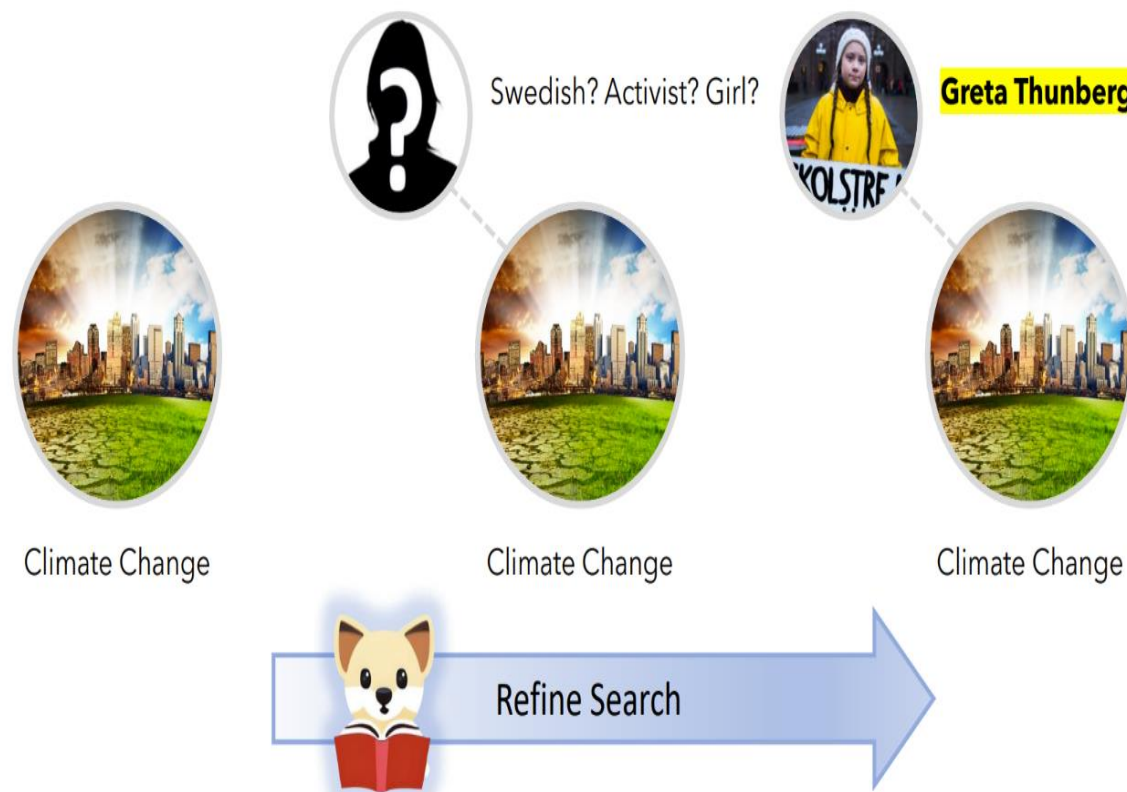


Search , and more Search (via Chatbot)

Solution Overview

Refine Search: Get closer to what you're really looking for

- Browser extension
- Query SMU Libraries' database with natural language
- OpenAI's GPT-4 processes users' inputs to capture their intention
- Intelligently convert users' queries into search parameters
- Search history captures users' past queries and allows search refinement
- Highlight and search text



Search , Search and more Search (via Chatbot)



Parrots GPT



How can I assist you?

Hi, I'm an SCIS student and I'm interested in learning more about human computer interactions. Could you give me some good resources so that I can learn more?



As an SCIS student interested in learning more about human-computer interaction (HCI), here are some recommended resources for you:



1. Book: "The Design of Everyday Things" by Don Norman - This classic book explores the principles of user-centered design and the psychology behind human-computer interaction.
2. Book: "Interaction Design: Beyond Human-Computer

Search , Search and more Search (via Chatbot)

Literature Review Generator

Please Input Variable X:

political connection

Please Input Variable Y:

accounting conservatism

Generate !

The impact of political connections on accounting conservatism has been a topic of interest in recent literature. Zhang, Nan, Liang, Qiaozhu an, Li, Huiying, and Wang, Xiao (2022) conducted a study to examine whether political connections lead to more or less opportunistic financial reporting. Using a difference-in-differences approach, the authors compared changes in financial reporting behavior before and after close elections involving politicians with influence over the SEC. The study utilized prior research studies, data on accounting fraud and restatements, and corporate governance variables. Surprisingly, the findings indicate that firms with increased political connections do not engage in more opportunistic reporting after close elections. In fact, these firms report less opportunistically, suggesting that political connections may enhance scrutiny by the SEC and mitigate opportunistic reporting. This study contributes to the understanding of the relationship between political connections and accounting conservatism, highlighting the potential benefits of political connections in promoting transparency and accountability in financial reporting.

The impact of political connections has been extensively studied in various contexts and countries. Palansky (2021) finds that politically connected firms in Bangladesh have reduced voluntary corporate social responsibility (CSR) disclosures, suggesting a decreased need for CSR as a legitimation strategy. Hung et al. (2018) examine the Egyptian context and find that firms with political connections through boards of directors and major shareholders are more likely to hire Big 4 audit firms and receive favorable audit opinions. Hoang et al. (2022) explore the association between political connections, political uncertainty, and audit fees in Pakistan, but their specific findings are not provided. Fengyan et al. (2022) focus on China and find that organizational relationship-based political connections positively affect private firms' access to bank loans, especially in regions with better institutional environments and industrial policy support. Xu and Yan (2020) investigate the impact of political connections on loan costs for privately held firms in China and find that connections to government officials lead to lower loan interest rates, with the effectiveness influenced by anti-corruption efforts. Zhao (2021) examines the relationship between political connections and financial reporting quality in Malaysian firms and finds that government-linked investments and politically connected directors are associated with lower-quality financial information. Chen et al. (2014) study politically connected firms in Taiwan and find that they receive preferential treatment in bank loans, particularly from government-owned banks and during presidential election years. Ahmad et al. (2022) analyze the relationship between corporate donations to political parties and firm performance in Czechia, finding that political connections are associated with superior reported profits. Tantavy and Moussa (2023) investigate the influence of political connections on the allocation of bailouts during the US financial crisis and find that being politically connected increased the relative bailout packages for connected firms. Akcigit et al. (2023) use a theoretical model to analyze the impact of political connections on revenue, investor income, and social welfare in China, finding that while connections increase managers' income, they also lead to higher moral hazard and a decrease in overall social welfare. Overall, these studies highlight the complex and varied effects of political connections on different aspects of firm behavior and performance in different countries and contexts.

No literature are in this part

No literature are in this part

To be fair our database/search vendors are thinking the same

Scopus AI: Change the way you view knowledge

Elsevier's new generative AI provides you with concise, trustworthy AI-based summaries of your queries.

Whether you are looking to get up to speed on a new field quickly or find cross-disciplinary collaborators to push your research to the next level. Scopus AI can help.

<https://beta.elsevier.com/products/scopus/scopus-ai>

Introducing **Dimensions AI Assistant**.

Launching with a closed beta-release, Dimensions is applying artificial intelligence to our vast database with the aim of speeding up data exploration and interpretation.

How will it work?

As part of the beta testing, users will be able to state a research interest and receive summaries generated by leveraging the Dimensions dataset and the large language models Dimensions General Sci-Bert and Open AI's ChatGPT.

This closed beta launch will collect feedback from our users and help us develop an even better data discovery experience, powered by cutting-edge artificial intelligence.

<https://www.dimensions.ai/products/all-products/dimensions-ai-assistant/>

Some academic search examples

Dimensions

Dimensions AI Assistant (beta)

is there an open access citation advantage?

Summary

This study found that gold OA physics journals showed a definite open access citation advantage, whereas other disciplines showed no difference or no open access citation advantage [1]. The comparative citation analysis indicated that APC-funded OA articles sustained their citation advantage in comparison with toll-access ones, and the citation advantage was apparently a sustainable phenomenon. Moreover, the number of APC-funded OA articles increased and exhibited citation advantages in almost all fields [2]. According to the results, the OA citation advantage that is observed in the whole sample still holds even for the highly similar OA and NOA papers, implying that the OA citation surplus is not an artefact of the OA and NOA papers' differences in their topics and, therefore, in their citation potential [3]. The results suggest that there is an open access citation advantage, however it varies depending on the field and discipline [4].

Dimensions.ai

SCISPACE

is there an open access citation advantag

Upload PDF

Login

Sign up

Insight from top 5 papers

English (en)

There is evidence of an open access citation advantage. Open access articles published in hybrid journals receive more citations than those published in gold open access journals [1]. Hybrid Gold open access articles in the field of Clinical Medicine have a greater number of citations compared to subscription articles [2]. Making open access copies of manuscripts available in self-archiving or "green" repositories leads to a positive citation effect [3]. However, the citation advantage of open access articles in Japanese institutional repositories is not confirmed [4].

Scispace

Elicit

is there an open access

is there an open access citation advantage?

SUMMARY OF TOP 4 PAPERS

BETA

The papers suggest that there is a citation advantage for open access articles, but the size of the advantage varies by discipline and the quality of the article. Ottaviani 2016 found that open access articles have a citation advantage of up to 19%, while Norris 2008 found that open access articles have a clear citation advantage

Elicit.org

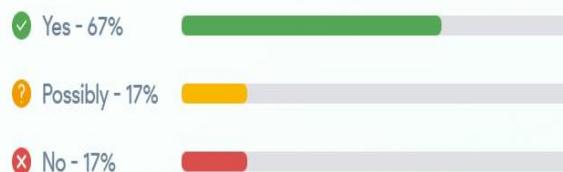
Summary

Top 7 papers analyzed

Some studies suggest an open access citation advantage (OACA) exists, with increased citation counts, end user uptake, and cross-discipline fertilization, while other studies find OACA only in subsets of their sample or in journals with low and medium impact factors.

Consensus Meter

6 papers analyzed



Consensus.ai



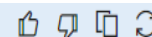
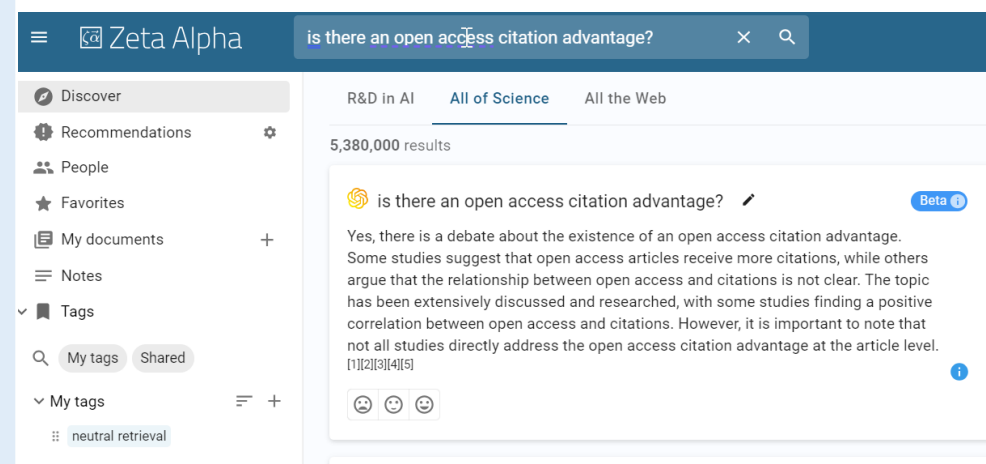
There is ongoing debate regarding whether there is an open access citation advantage. Proponents of open access argue that open access articles are more quickly recognized and cited ([Eysenbach, 2006](#)). They suggest that freely available articles are more likely to be read and cited compared to subscription-only articles ([Davis et al., 2008](#)). Some studies have found weak but statistically significant correlations between the number of article downloads and future citations ([Davis et al., 2008](#)). Additionally, research has shown that open access articles receive more citations in certain fields, such as medicine and radiology ([Alkhawtani et al., 2019](#)).

However, critics of open access dispute the existence of a citation advantage. They argue that there is no evidence to support the claim that open access articles are cited more frequently ([Eysenbach, 2006](#)). Some traditional publishers argue that the established system of scientific publishing already provides access to scientists and the public, minimizing the advantage of open access ([Eysenbach, 2006](#)).

A systematic review of studies comparing citations to open access and non-open access articles found contradictory and inconclusive results ([Langham-Putrow et al., 2021](#)). While some studies have reported a citation advantage for open access articles, others have found no general open access citation advantage ([Norris et al., 2008](#)). It is important to note that the presence of an open access citation advantage may vary across disciplines and journals ([Clements, 2017](#)).

In addition to traditional citation metrics, social media attention has emerged as a potential indicator of impact. Research has shown a relationship between academic impact (measured by citations) and social impact (measured by social media attention) for open access books ([Wei & Chakoli, 2020](#)). Open access books have been found to boost knowledge distribution on platforms like Twitter and Mendeley ([Wei & Chakoli, 2020](#)).

Scite.ai assistant

The screenshot shows the Zeta Alpha search interface. The search bar at the top contains the query "is there an open access citation advantage?". Below the search bar, there are tabs for "R&D in AI", "All of Science" (selected), and "All the Web". The results section shows "5,380,000 results". A snippet of a result is displayed, starting with "is there an open access citation advantage?". The snippet text reads: "Yes, there is a debate about the existence of an open access citation advantage. Some studies suggest that open access articles receive more citations, while others argue that the relationship between open access and citations is not clear. The topic has been extensively discussed and researched, with some studies finding a positive correlation between open access and citations. However, it is important to note that not all studies directly address the open access citation advantage at the article level. [1][2][3][4][5]". There are also three small circular icons (neutral, happy, sad) below the snippet.

Zeta alpha

is there an open access citation advantage?

✓ Searching for: **open access citation advantage**

✓ Generating answers for you...

The Open Access Citation Advantage (OACA) is a concept that suggests scholars tend to cite academic journals with open access (OA)—journals that make their full text available on the Internet without charge—in preference to toll-access publications ¹.

However, the existence of an open access citation advantage has been a topic of much discussion over the last two decades. While there has been substantial research to address this question, findings have been contradictory and inconclusive ². A systematic review identified 134 papers on this topic and found that 47.8% of them found evidence for OACA, 27.6% found that it did not exist, 23.9% found OACA in subsets of their samples and only 1 study (0.8%) did not find any OACA ³.

Therefore, while some studies suggest an advantage, the evidence is not consistent across all research. It's also important to note that most studies are observational and do not properly control confounding effects ³. For example, authors might be more motivated to pay Article Processing Charges (APCs) to make their best works open access, which could influence citation comparisons between open access and paywalled articles ³.

Learn more: [1. bing.com](#) [2. journals.plos.org](#) [3. library.smu.edu.sg](#) [+4 more](#)

1 of 30 ●

Bing Chat

is there an open access citation advantage?

⚡ Quick Search

Is the open access citation advantage real? ...

 plos • 1

Open access citation advantage - Wikipedia

 wikipedia • 2

Open access articles attract more citations - ...

 frontiersin • 3

Is the open access citation advantage real? ...

 nih • 4

Rethinking the open access citation...

 wiley • 5

Do open access journal articles experience a...

 springer • 6

≡ Answer

There is no clear consensus on whether an open access citation advantage (OACA) exists. While some studies have found evidence of an OACA, others have found no significant difference in citation rates between open access and subscription-based articles. A systematic review of studies on the topic found that of the studies with low risk of bias, one found an OACA existed, one found it did not, and one found it occurred in subsets ¹. Another analysis of highly cited studies found that 43 out of 58 studies confirmed the existence of an OACA ³. However, a more recent study found that the extra visibility brought by open access is likely from people who do not engage with the articles in terms of citations ⁵. Overall, the evidence is mixed and inconclusive, and more research is needed to determine whether an OACA exists.

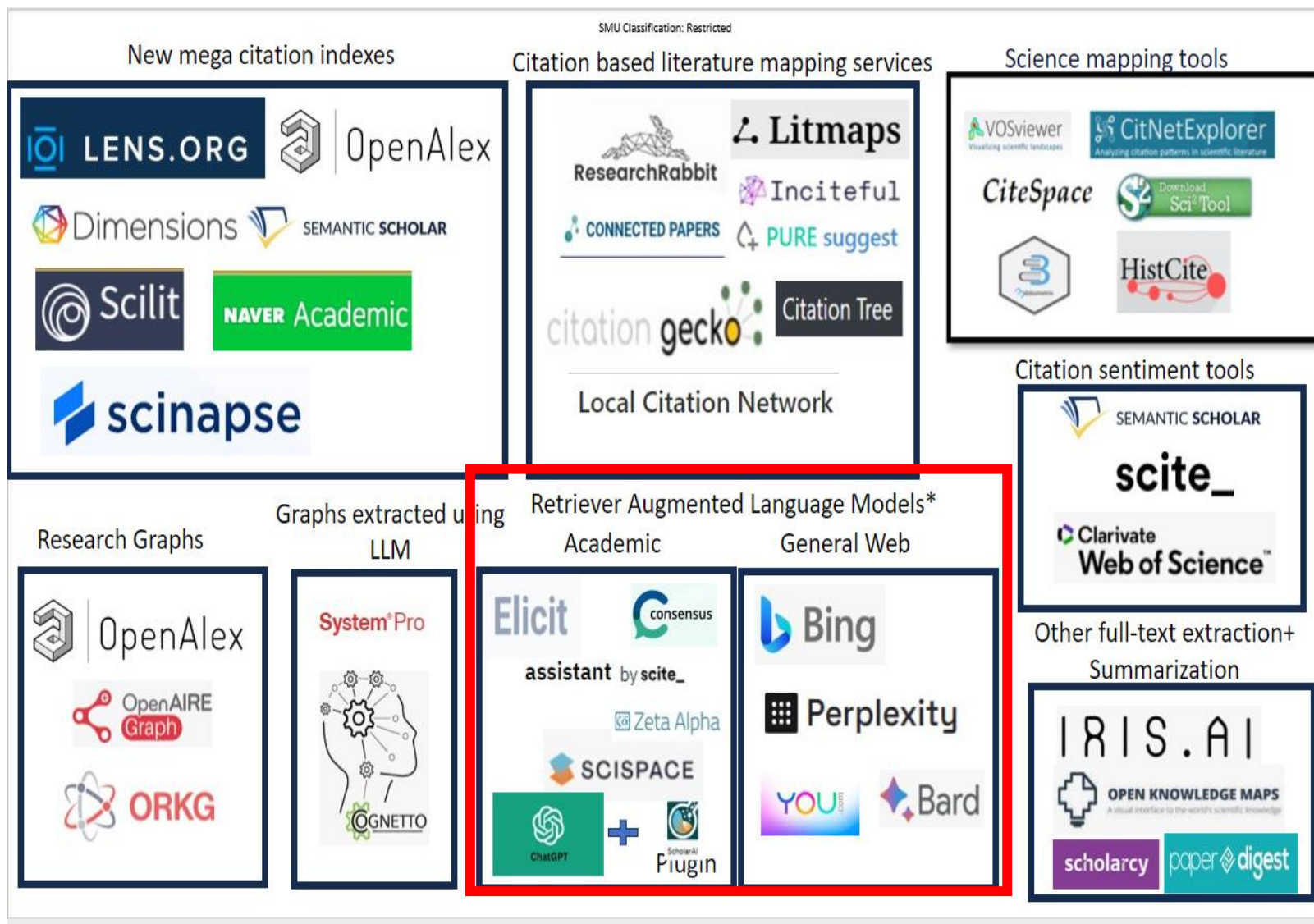
Perplexity.ai

Predicted trends (within 3 years)

- Most/many academic search engines will include generative AI/LLM (Large language models) – **VERY HIGH (95%)**
- Users will get more and more used to natural language search style and generally trust direct generated answers – **High (80%)**
- Reasonably decent chatbots will be deployed in libraries for first tier support – **Moderately High 75%**
- Systems will exist that do a generally adequate job of writing a literature review (2 papers long) – **Medium (50%)**

How generative AI/LLMs are affecting search engines

AI + Search?



The confusing world of academic discovery

What do we mean by "AI"?

- These days AI = Large Language Models (Transformer architecture)
- OpenAI's – GPT3.5-Turbo, GPT4
- Anthropic - Claude 2
- Google/Deep Mind – BERT*, T5*, **Chinchilla**,
•LaMDA, PaLM2
- Meta's – Llama2, Falcon 180B (Open Source**)
- AI2 - OLMo (Academic Open Language Model)

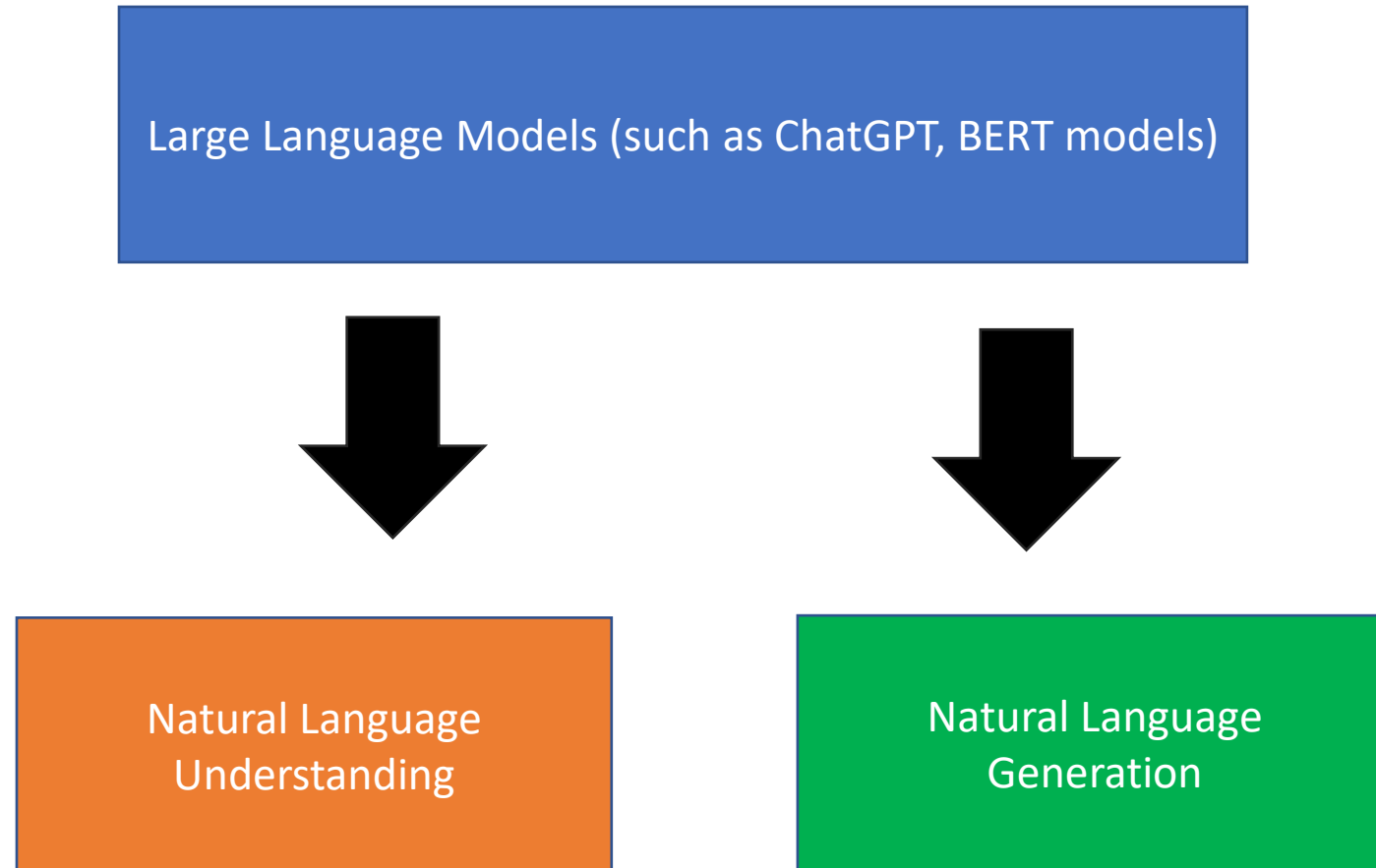
* Most of the list are decoder only models, BERT is encoder model
and T5 is Encoder decoder model

** Some restrictions

Hype or reality?

"GPT-4 exhibits human-level performance on the majority of these professional and academic exams. Notably, it passes a simulated version of the Uniform Bar Examination with a score in the top 10% of test takers."

Simulated exams	GPT-4 estimated percentile	GPT-4 (no vision) estimated percentile	GPT-3.5 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT) ¹	298/400 ~90th	298/400 ~90th	213/400 ~10th
LSAT	163 ~88th	161 ~83rd	149 ~40th
SAT Evidence-Based Reading & Writing	710/800 ~93rd	710/800 ~93rd	670/800 ~87th
SAT Math	700/800 ~89th	690/800 ~89th	590/800 ~70th
Graduate Record Examination (GRE) Quantitative	163/170 ~80th	157/170 ~62nd	147/170 ~25th
Graduate Record Examination (GRE) Verbal	169/170 ~99th	165/170 ~96th	154/170 ~63rd
Graduate Record Examination (GRE) Writing	4/6 ~54th	4/6 ~54th	4/6 ~54th
USABO Semifinal Exam 2020	87/150 99th–100th	87/150 99th–100th	43/150 31st–33rd
USNCO Local Section Exam 2022	36/60	38/60	24/60
Medical Knowledge Self-Assessment Program	75%	75%	53%
Codeforces Rating	392 below 5th	392 below 5th	260 below 5th
AP Art History	5 86th–100th	5 86th–100th	5 86th–100th
AP Biology	5 85th–100th	5 85th–100th	4 62nd–85th



"Given the breadth and depth of GPT-4's capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system." - [Sparks of Artificial General Intelligence: Early experiments with GPT-4](#)

What are language models?

A language model is a statistical model that allows you to predict the probability of the next word in a sequence of words.

The best thing about AI is its ability to

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

Terminology

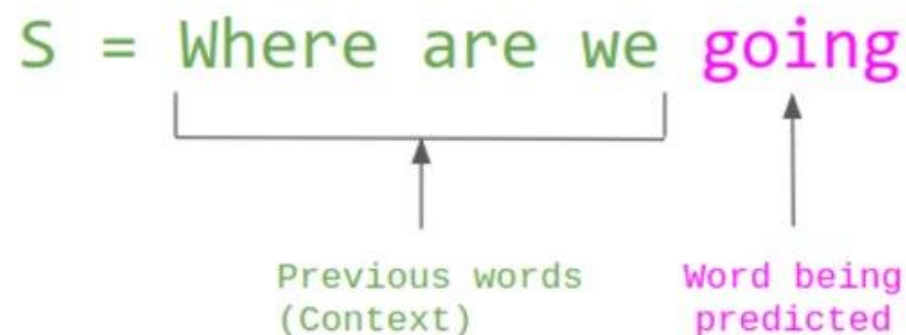
OpenAI's **ChatGPT** is a Large Language Model based on transformer architecture

A **large language model (LLM)** is a type of machine learning model trained on vast amounts of text data. It learns to understand and generate human-like text by *predicting the next word in a sentence, given the words that came before it*. Large language models often use the transformer architecture to process and generate text, thanks to its powerful capabilities.

A **transformer model** is a type of *artificial neural network architecture* that is designed to process and generate sequences of data, like text, in a highly parallel and efficient way.

OpenAI's **ChatGPT** is a Large Language Model based on transformer architecture

Use in smartphone keyboards, autocomplete/correct



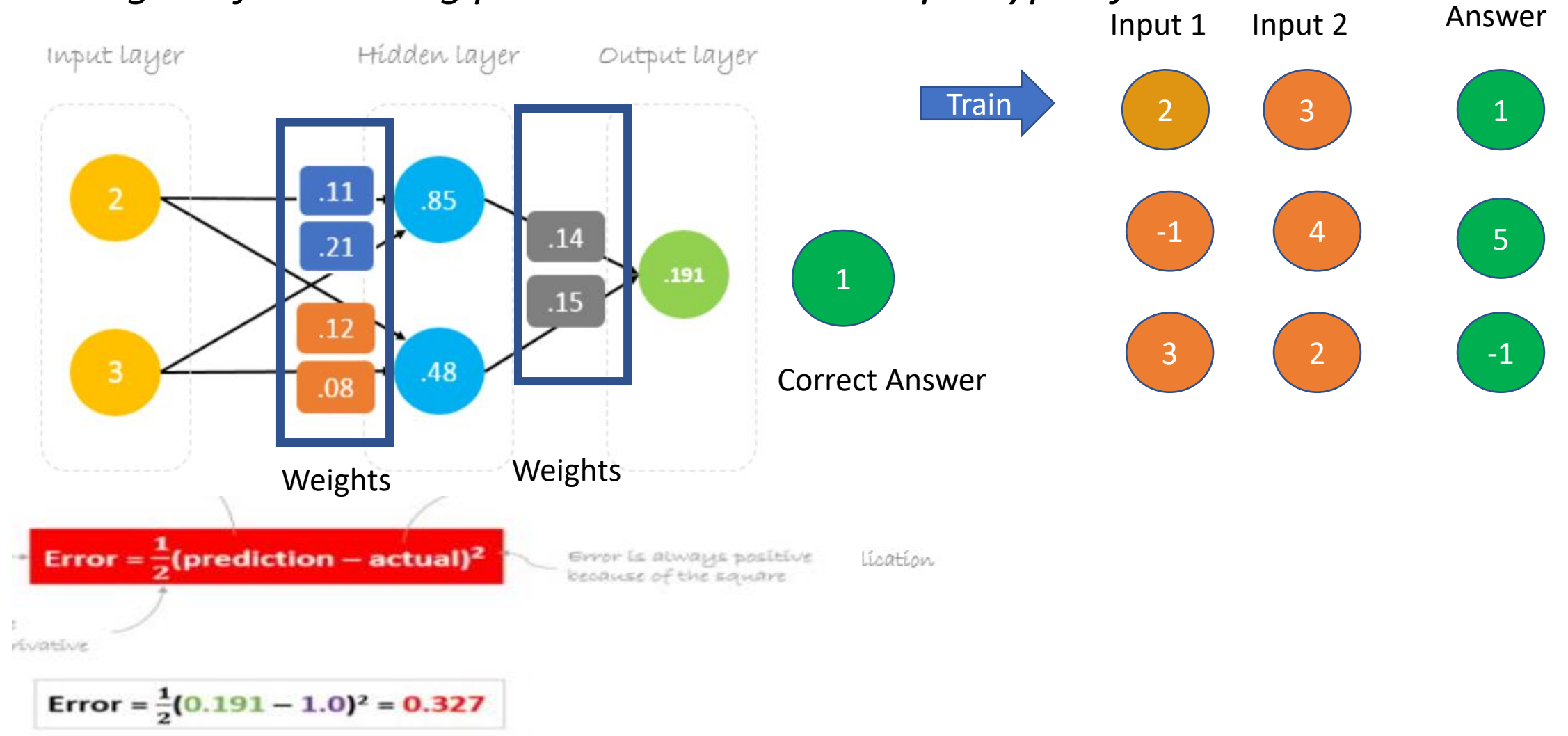
$$P(S) = P(\text{Where}) \times P(\text{are} \mid \text{Where}) \times P(\text{we} \mid \text{Where are}) \times P(\text{going} \mid \text{Where are we})$$

What is a neural network and why do we need it?

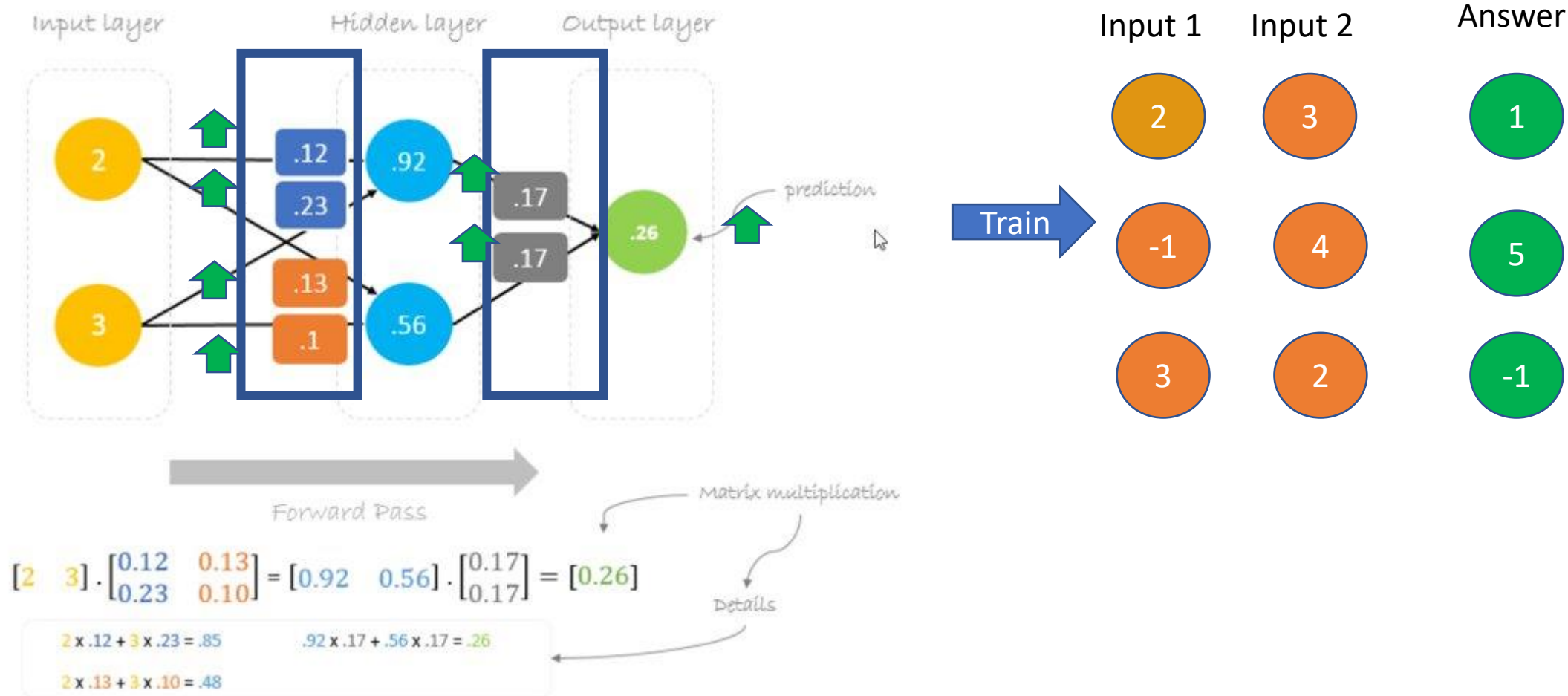
A **transformer model** is a type of *artificial neural network architecture* that is designed to process and generate sequences of data, like text, in a highly parallel and efficient way.

*An artificial neural network is designed to mimic how human brains work and enables creation of systems that can **teach itself how to do a task** (machine learning).*

Transformer model is a type of *artificial neural network architecture*. *Neural nets are good for learning patterns – Here's a simple type of neural net*



Weights in neural net are adjusted using back propagation using gradient descent



Language Models are trained in generation do these tasks

Next Token Prediction (Causal Language Modelling)

Predicting the next word in a _____

Masked Language Modelling

Predicting the _____ that should go into the sentence

Other tasks

- **Next Sentence Prediction**
- **Sentence Boundary Objective**
- **Etc**

ChatGPT is Large Language Models(LLMs) - based on Transformers architecture paper (2017)

Encoder models (**bidirectional**)

- BERT
- RoBERTa
- DistilBERT

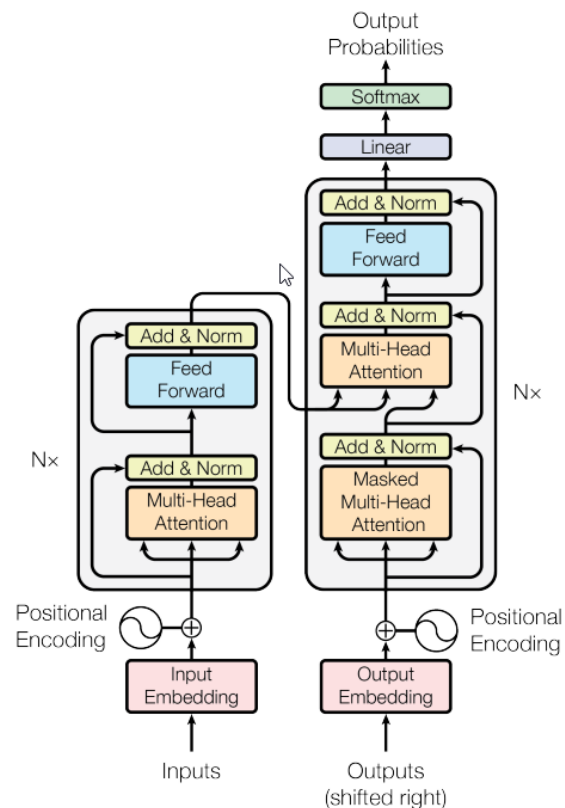
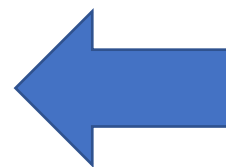
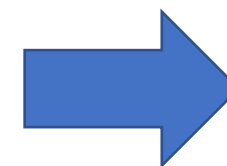


Figure 1: The Transformer - model architecture.

Decoder models (**autoregressive**)

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

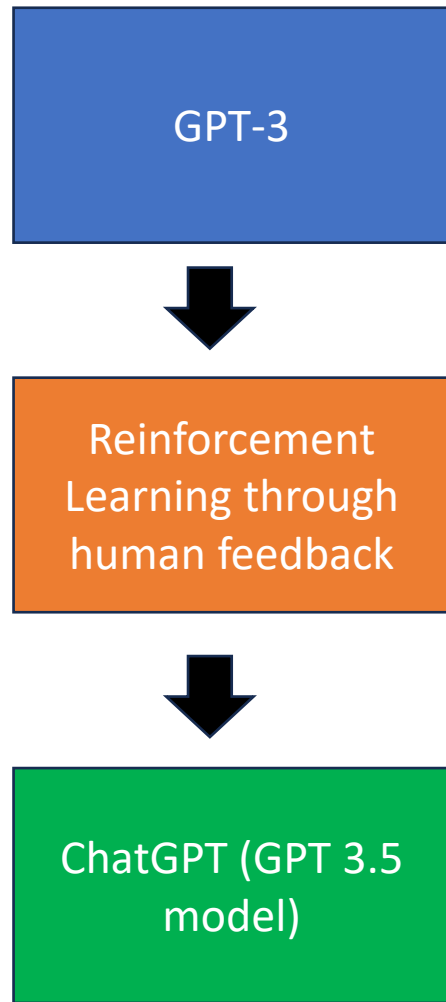


[Attention is all you need](#)

Final Step – GPT3 to Instruction Aligned (GPT3.5 eg ChatGPT)

- GPT models are basically trained to generate text based on associations it learnt from past training data .
- There is no certainty it's text output will
 - follow instructions you input
 - Not make up things/tell the truth (factuality) -
 - Not be output biased or dangerous information or advice.
 - Be conversational

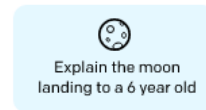
Final Step – GPT3 to Instruction Aligned via RLHF



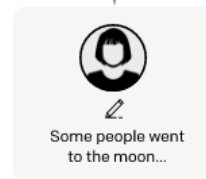
Step 1

Collect demonstration data, and train a supervised policy.

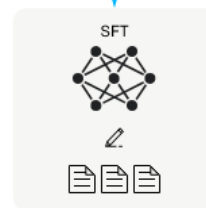
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



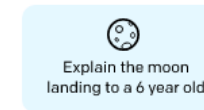
This data is used to fine-tune GPT-3 with supervised learning.



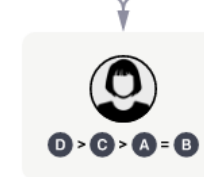
Step 2

Collect comparison data, and train a reward model.

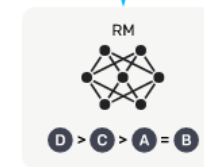
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using reinforcement learning.

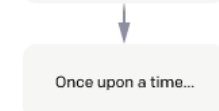
A new prompt is sampled from the dataset.



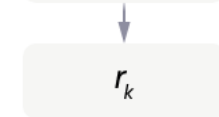
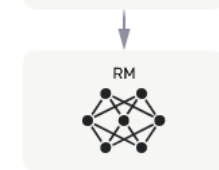
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



<https://openai.com/research/instruction-following>

**Why hallucinations happen
when using just Language
models like ChatGPT**



who is aaron tay the academic librarian in Singapore



Aaron Tay is an academic librarian based in Singapore, known for his work in the field of library and information science. He has been associated with the National University of Singapore (NUS) Libraries and has contributed significantly to the library profession through his research, publications, and active engagement in the library community.

As of my last knowledge update in September 2021, Aaron Tay was working as the Assistant Director for Library Analytics and Metrics at the National University of Singapore (NUS) Libraries. He is well-known for his expertise in library analytics, data visualization, and the application of technology in libraries. Aaron Tay has also been active on social media and maintains a blog where he shares insights and information related to library science and technology.

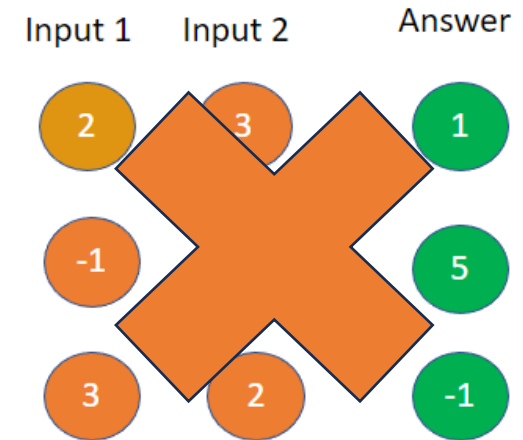
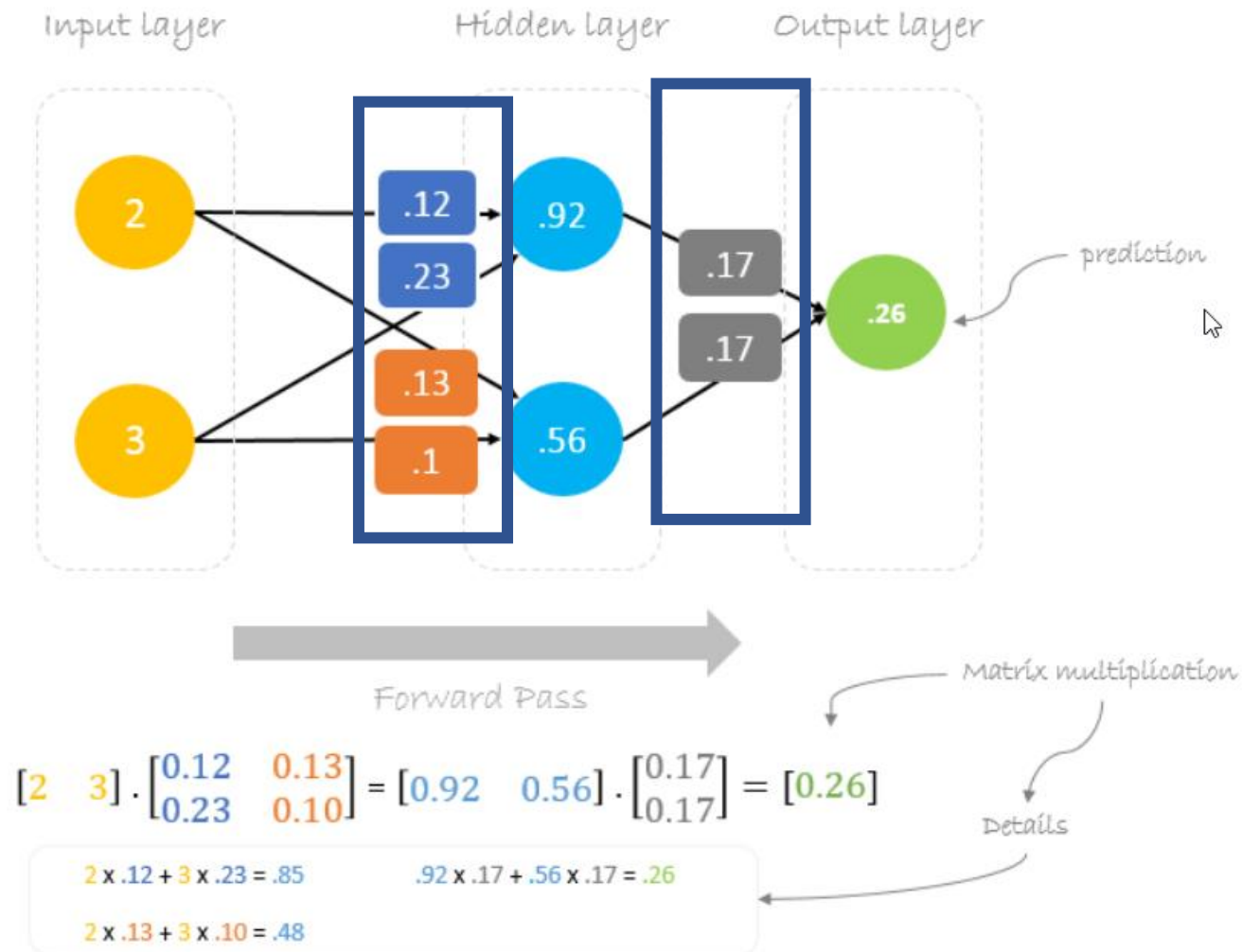
For the most current information about Aaron Tay's professional activities and affiliations, I recommend visiting the official website of the NUS Libraries or checking his professional profiles on platforms like LinkedIn or Twitter, where he often shares updates about his work.

How databases/search engines like Google store text – inverted index

Term	Doc 1	Doc 2	Doc 3
Aaron	X	X	
Tay	X	X	X
Singapore	X	X	
Management	X		x
University	X	X	
Librarian		X	x

How databases or search engine index store text – inverted index

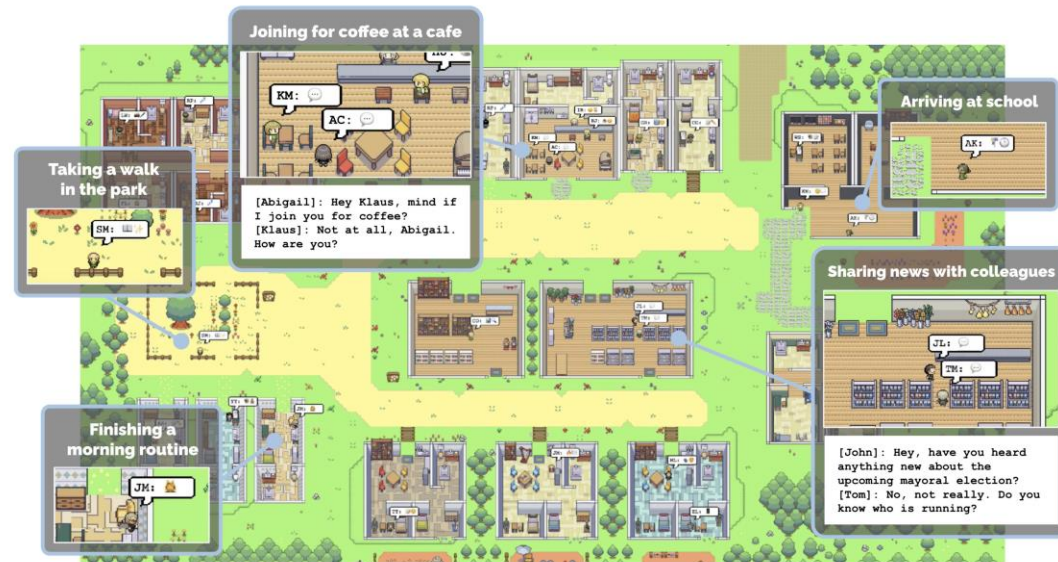
Why does it make things up? It is not a database! It stores weights



All the training data is discarded

LLM + other tech to improve Science and academia

- Text sentiment
- Text extraction
- Error checking
- Data cleaning and anonymization
- Simulations
- Hypothesis Generation
- Proof generation
- Many ,Many, more



Generative Agents: Interactive Simulacra of Human Behavior

How do LLM's improve search (2023)

Current "standard" (2023) use

- Improved relevancy (through contextual embeddings)
- **Generation of direct answers (paragraph or two)**
- Extraction of information from papers (abstract and full-text) to enhance search engine result

Things people working on

- Semi automation of systematic reviews
- Intergration with Knowledge Graph
- **Literature review generation** (holy grail)

1. Improved relevancy through contextual embeddings

Lexical/keyword search vs Semantic/Neural search

Search query - how to find seminal papers?

Google Scholar

Articles About 731,000 results (0.07 sec)

Any time
 Since 2023
 Since 2022
 Since 2019
 Custom range...

Sort by relevance
 Sort by date

Any type
 Review articles

☐ include patents
☒ include citations
☒ Create alert

Control theory: Twenty-five seminal papers, 1931-1981
[T Basar](#), [T Başar](#) - 2001 - [iee.org](#)
 ... This product is added to the list of offerings that assist members as they **find** jobs and ... You can also read first-hand stories from PES Scholars about why they picked their major, **how** they'...

☆ Save ⓘ Cite Cited by 65 Related articles All 3 versions ⓘ

A model of sales
[HR Varian](#) - [The American economic review](#), 1980 - JSTOR
 ... **how** such price dispersion can persist in markets where at least some consumers behave in a rational manner. Starting with the **seminal paper** ... able so we can **find** an arbitrarily small e ...

☆ Save ⓘ Cite Cited by 2835 Related articles All 5 versions ⓘ

Exploration and exploitation in innovation: Reframing the interpretation
[Y Li](#), [W Vanhaverbeke](#) ... - [Creativity and innovation](#) ..., 2008 - [Wiley Online Library](#)
 ... March's **seminal** article in 1991. However, in reviewing the extant literature we **find** different ... Second, **how** can we set up a framework that reconciles these differences and reduces the ...

☆ Save ⓘ Cite Cited by 401 Related articles All 13 versions ⓘ

[Google Scholar](#)

Elicit

how do you find seminal papers?

SUMMARY OF TOP 4 PAPERS **BETA**

The papers suggest several methods for finding seminal papers. Hernández (2020) used data mining techniques to detect seminal papers in three different areas of knowledge, finding that seminal articles can be different depending on whether Google Scholar or Scopus is used. Runelöv (2015) investigated the applicability of network analysis to the problem of finding seminal publications, finding that the backbone graph provides a way to possibly discover seminal publications with low citation count. Haunschild (2020) presented a strategy to discover the most important papers of a


Add information about all papers
 Abstract summary ✓
 Intervention
 Outcomes measured

Search for paper information

Paper title

☆ **Findings Seminal Papers Using Data Mining Techniques**
 ⓘ A. Hernández, Debrayan Bravo Hidalgo
 2020 0 Citations DOI ⓘ

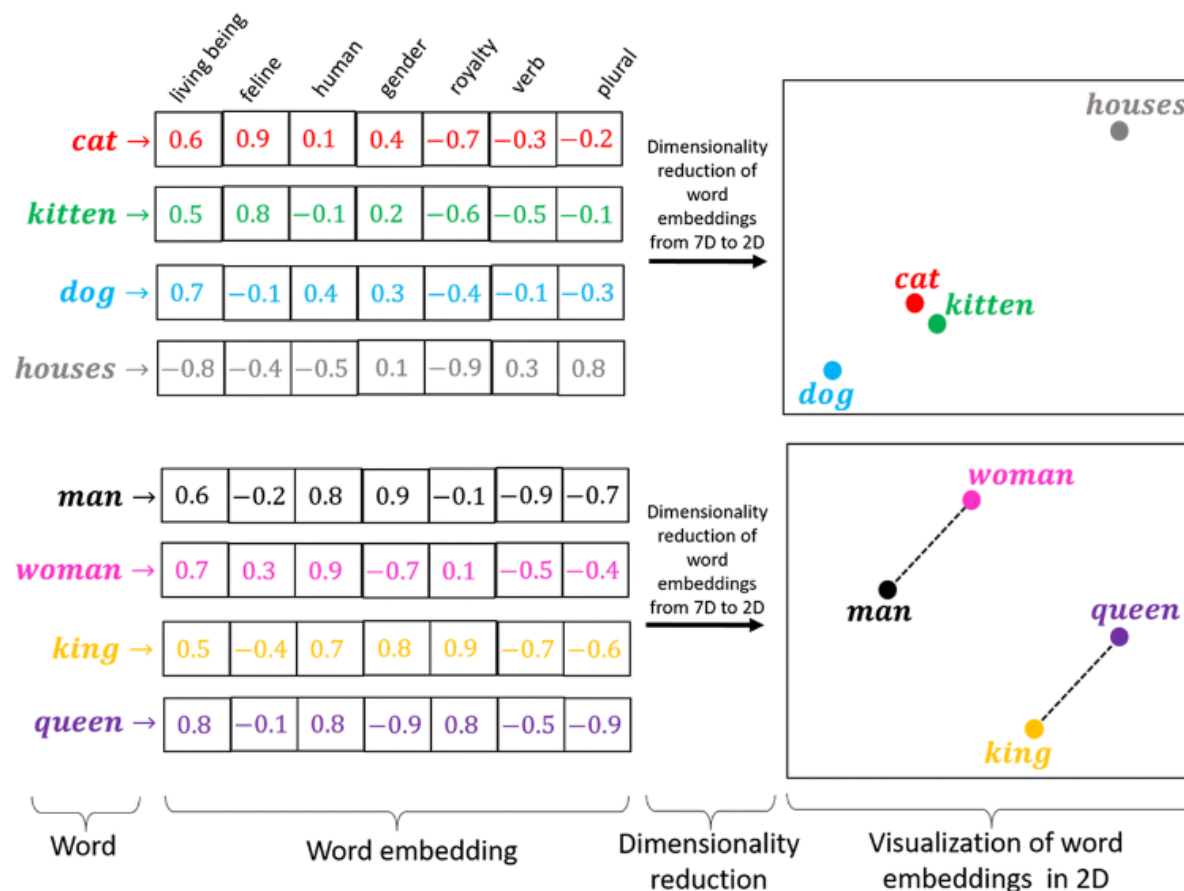
☆ **Finding seminal scientific publications with graph mining**
 ⓘ Martin Runelöv
 2015 0 Citations

☆ **Discovering seminal works with marker papers**
 ⓘ Robin Haunschild, Werner Marx
 ⓘ [Scientometrics](#) 
 2020 10 Citations DOI ⓘ

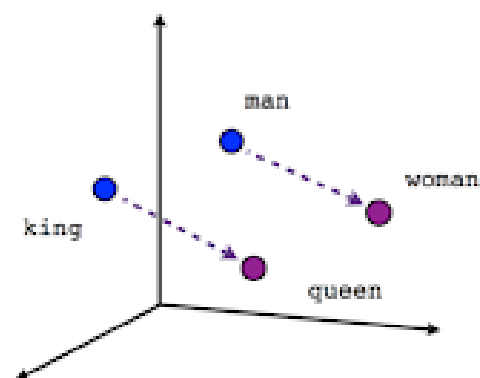
☆ **Collecting the Seminal Scientific Abstracts with Topic Modelling, Snowball Sampling and Citation Analysis**
 ⓘ H. Dobrovolskyi, Nataliya Keberle
 ⓘ [International Conference on Information and Communication Tech...](#)
 2018 4 Citations

[Elicit.org](#)

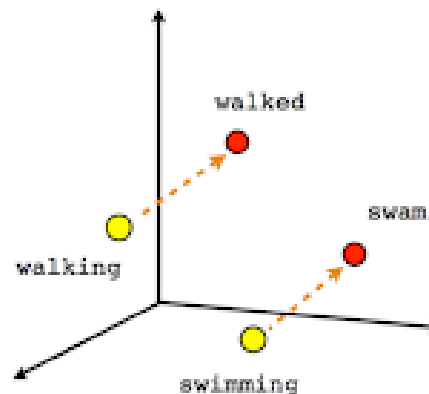
Embeddings – converting words into semantic meaning



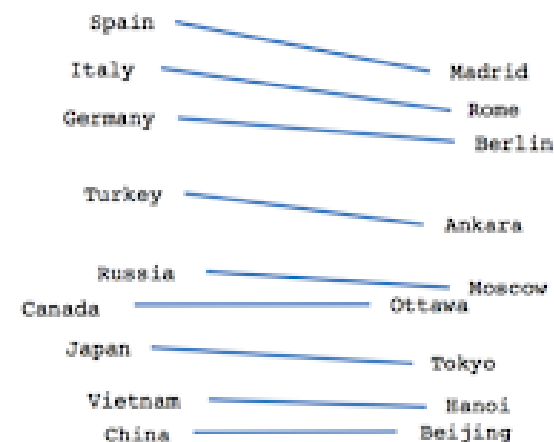
Word2Vec (2013)



Male-Female



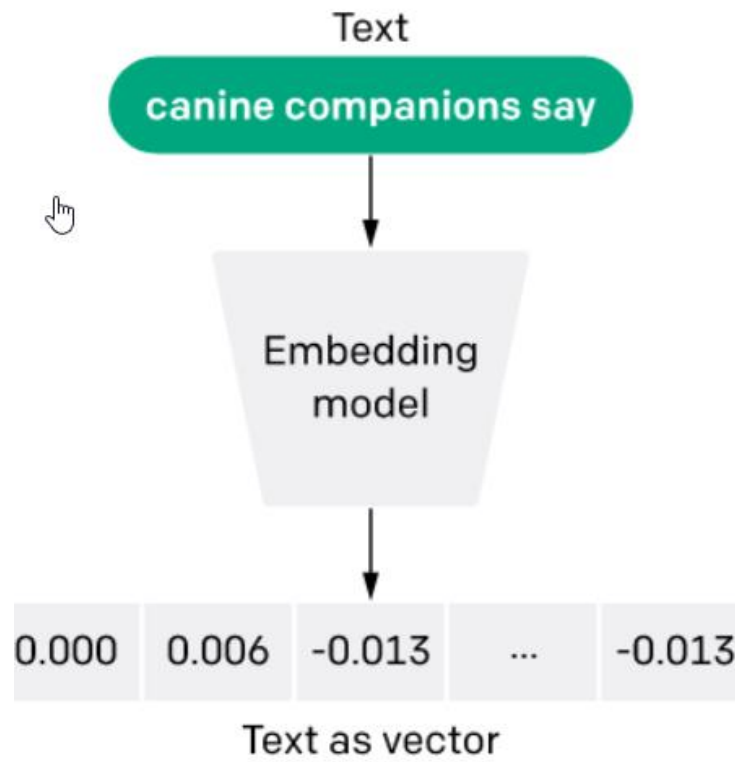
Verb tense



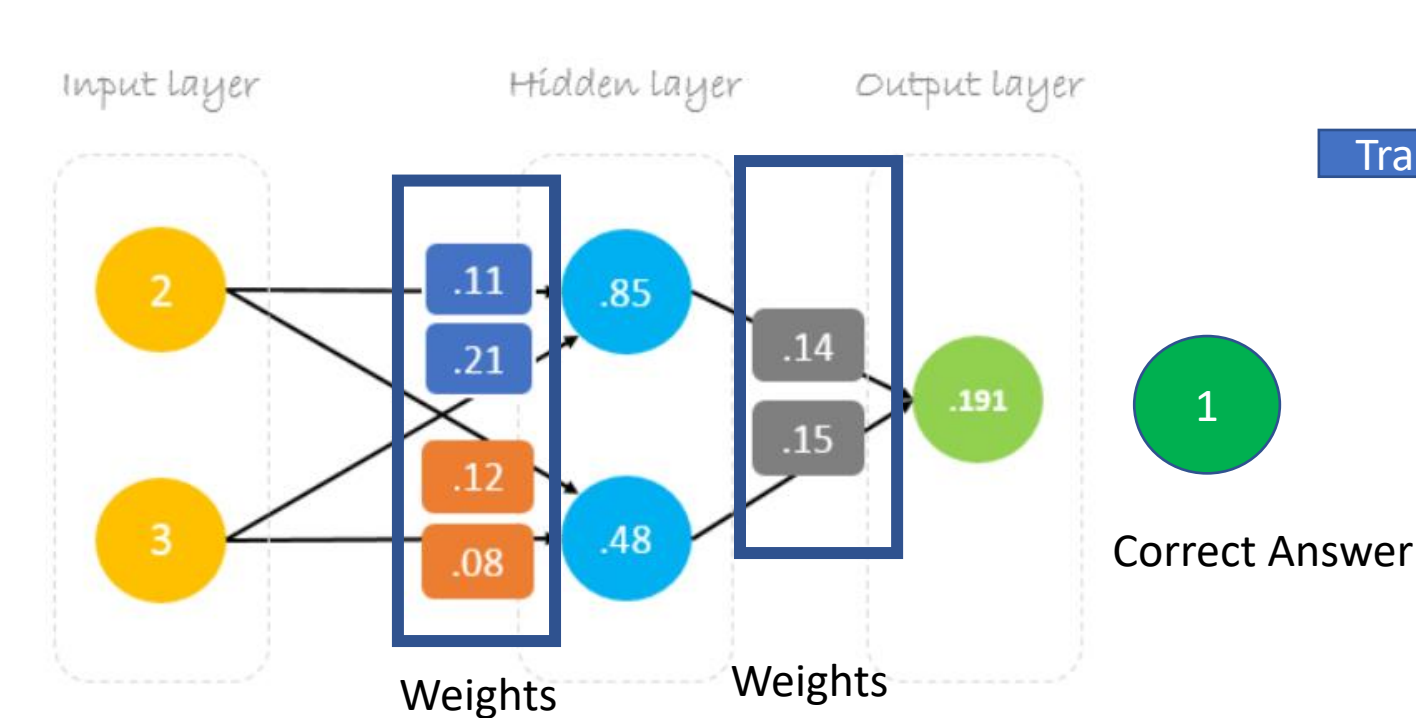
Country-Capital

MAN-KING+WOMAN = QUEEN

[Word2Vec \(2013\)](#)



Remember the neural network from earlier?



Input 1	Input 2	Answer
2	3	1
-1	4	5
3	2	-1

$$\text{Error} = \frac{1}{2}(\text{prediction} - \text{actual})^2$$

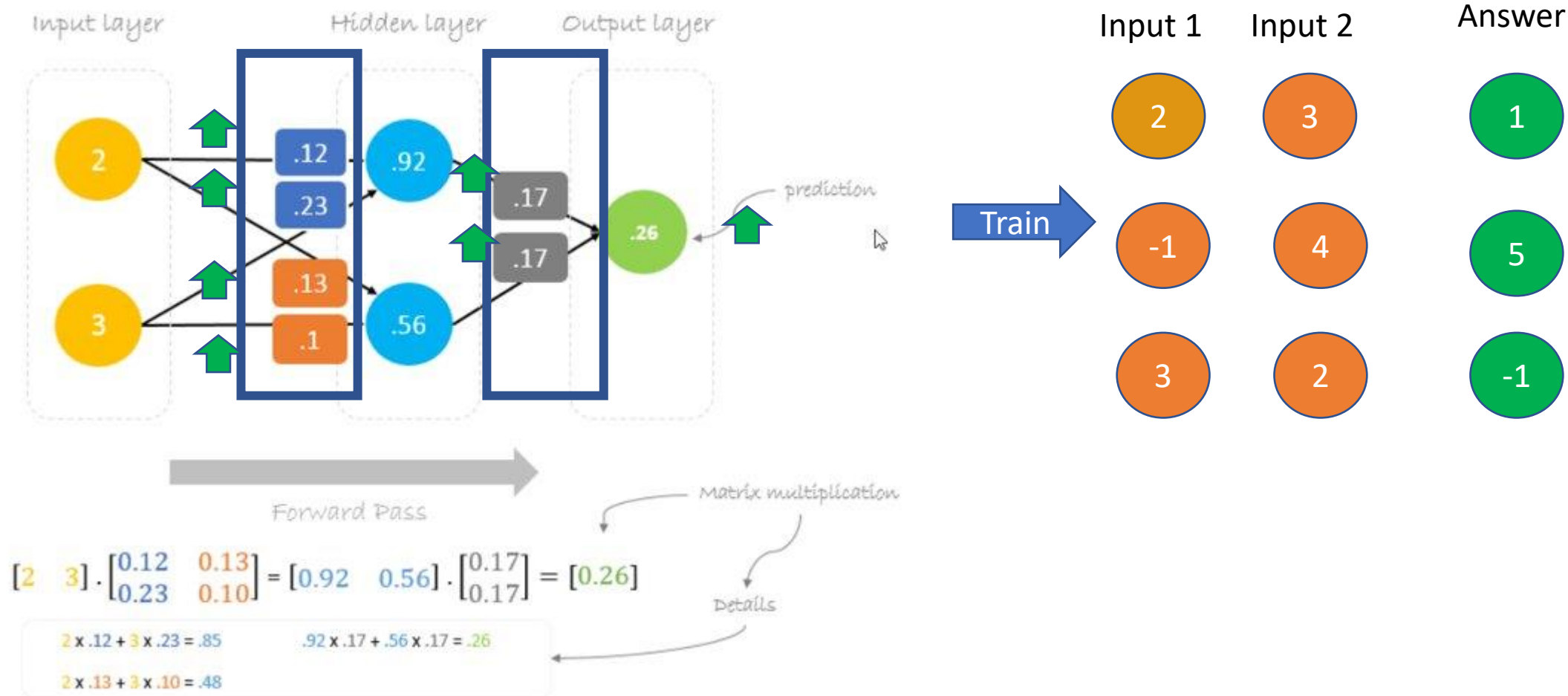
Error is always positive because of the square

derivative

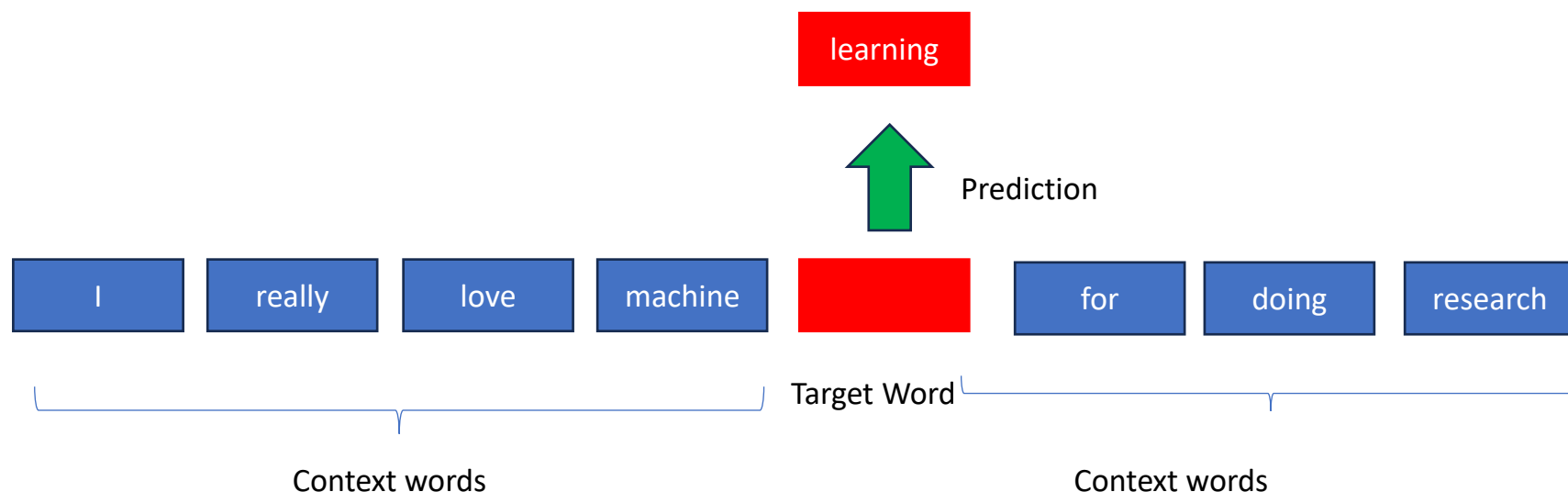
location

$$\text{Error} = \frac{1}{2}(0.191 - 1.0)^2 = 0.327$$

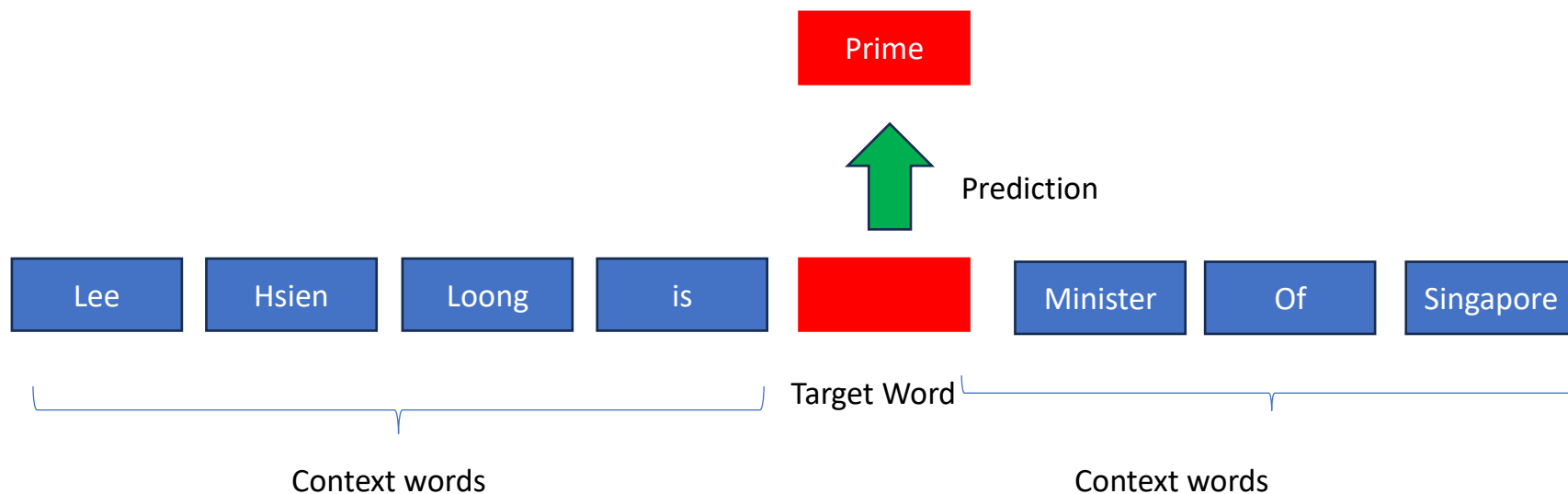
Weights in neural net are adjusted using back propagation using gradient descent



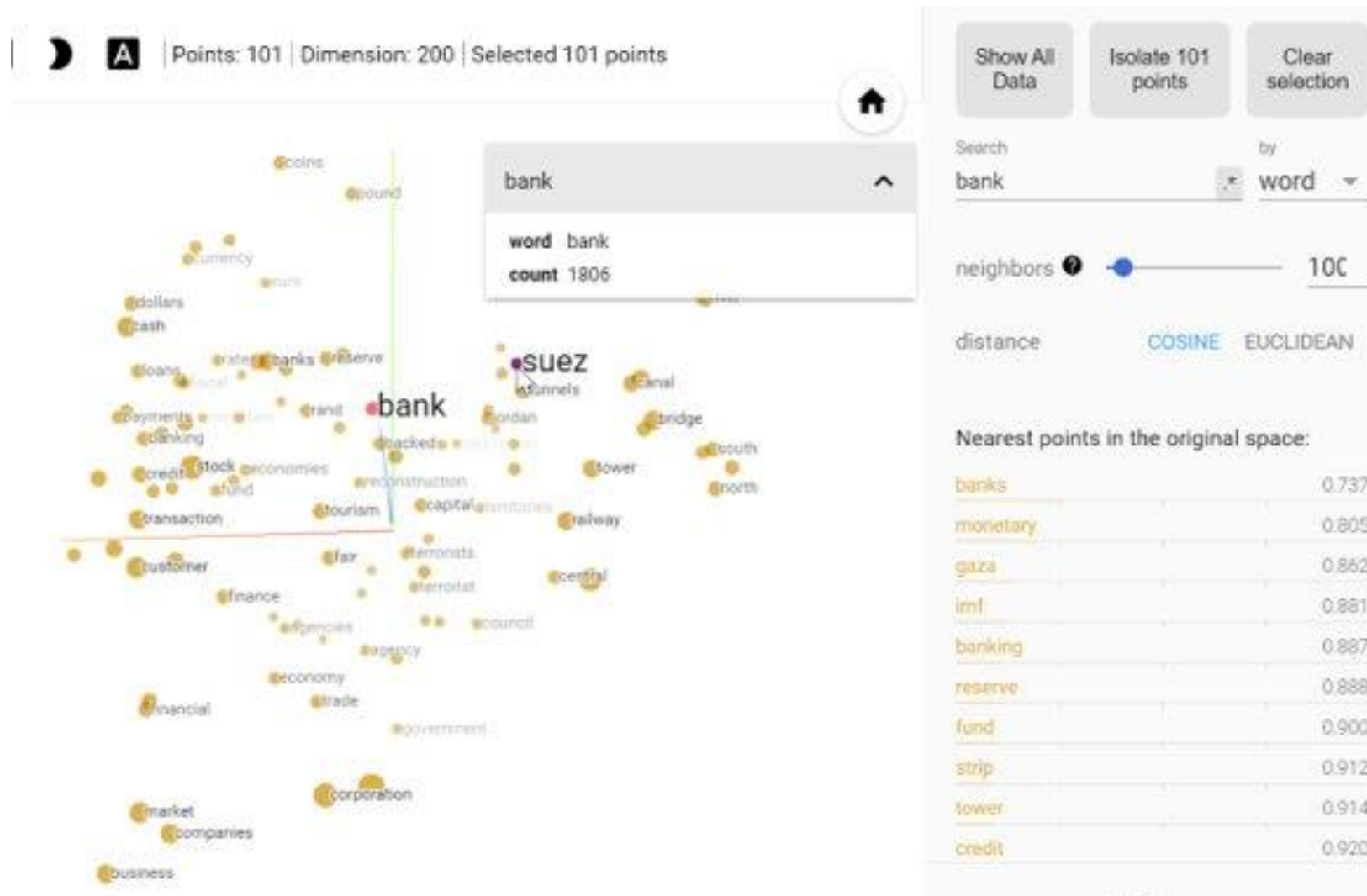
Continuous bag-of-words (CBOW) predicts target words from context words



Based on statistics patterns/distributions not fact



Next advancement – solving the problem of Homonyms – River Bank vs Financial Bank



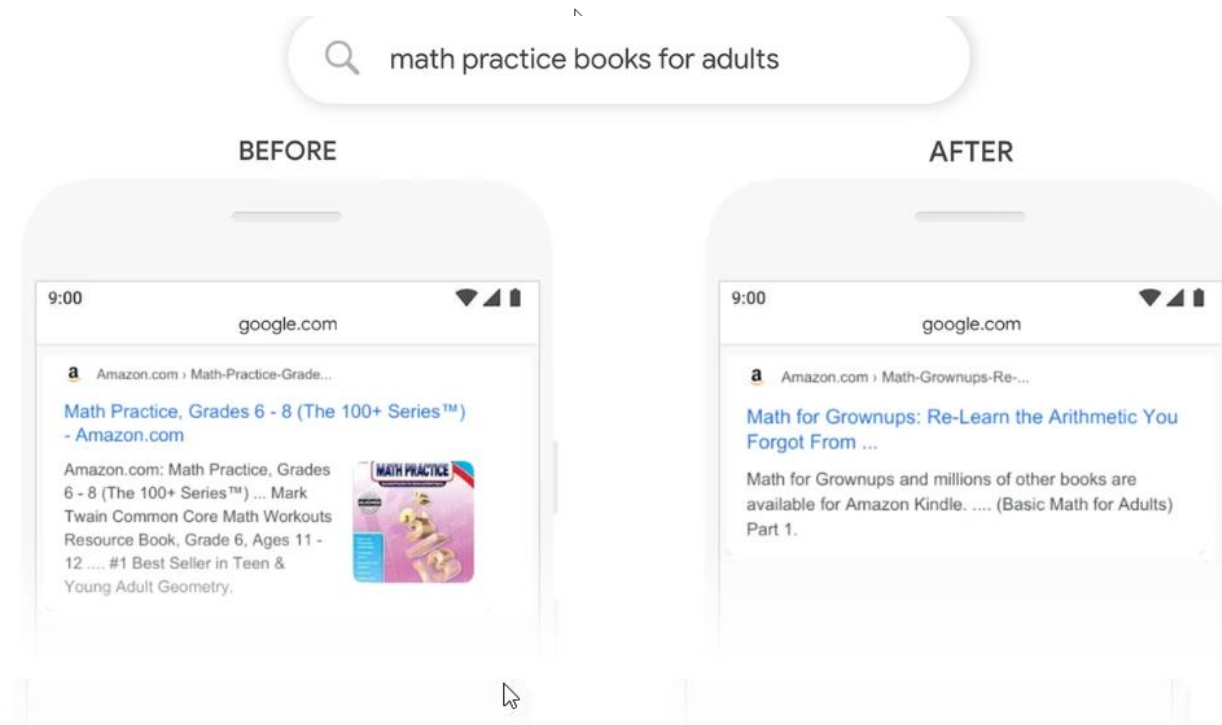
Transformers use self-attention to further refine meaning of words

Similarities	New words	Attention step	Normalization step																									
<p>The bank of the river</p> <table><tr><th></th><th>the</th><th>bank</th><th>of</th><th>river</th></tr><tr><th>the</th><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><th>bank</th><td>0</td><td>1</td><td>0</td><td>0.11</td></tr><tr><th>of</th><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><th>river</th><td>0</td><td>0.11</td><td>0</td><td>1</td></tr></table>		the	bank	of	river	the	1	0	0	0	bank	0	1	0	0.11	of	0	0	1	0	river	0	0.11	0	1	<p>the1 bank1 of1 river1</p>	<p>the bank +0.11 river of river +0.11 bank</p>	<p>the +0.9 bank +0.1 river of +0.9 river +0.1 bank</p>
	the	bank	of	river																								
the	1	0	0	0																								
bank	0	1	0	0.11																								
of	0	0	1	0																								
river	0	0.11	0	1																								
<p>Money in the bank</p> <table><tr><th></th><th>money</th><th>in</th><th>the</th><th>bank</th></tr><tr><th>money</th><td>1</td><td>0</td><td>0</td><td>0.25</td></tr><tr><th>in</th><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><th>the</th><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><th>bank</th><td>0.25</td><td>0</td><td>0</td><td>1</td></tr></table>		money	in	the	bank	money	1	0	0	0.25	in	0	1	0	0	the	0	0	1	0	bank	0.25	0	0	1	<p>money2 in2 the2 bank2</p>	<p>money +0.25 bank in the bank +0.25 money</p>	<p>+0.8 money +0.2 bank in the +0.8 bank +0.2 money</p>
	money	in	the	bank																								
money	1	0	0	0.25																								
in	0	1	0	0																								
the	0	0	1	0																								
bank	0.25	0	0	1																								

Understanding searches better than ever before

Oct 25, 2019 · 5 min read

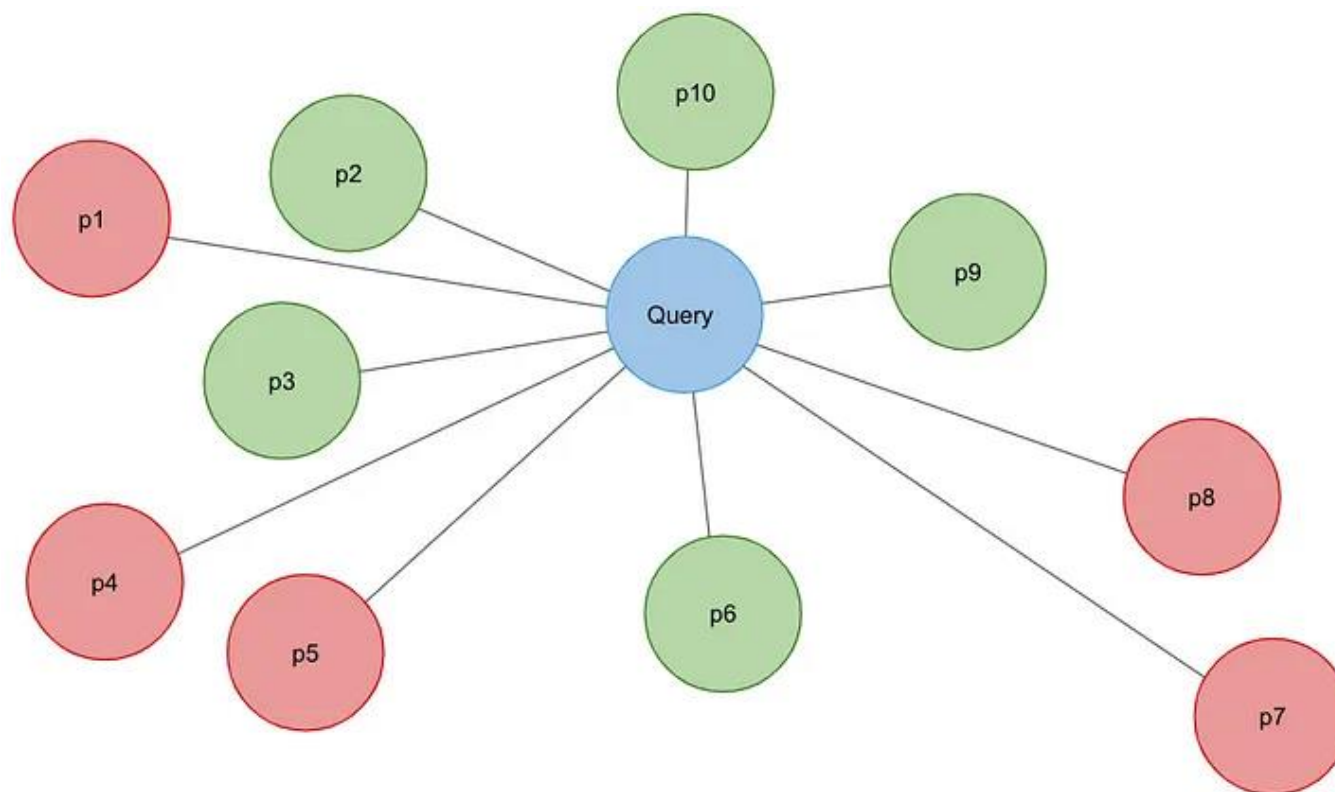
Google and other search engines have quietly used LLMs like BERT models since 2019!



<https://blog.google/products/search/search-language-understanding-bert/>

Implications

Harder and harder to “explain” search results



Implications

How should one search in such search engines?

1. **Keyword style** - Open access citation advantage?
2. **Natural language style** - Is there an Open Access Citation advantage (natural language)
3. **Prompt engineering style**** – You are a top researcher in the subject of Scholarly communication. Write a 500 word essay on the evidence around Open Access Citation advantage with references

** What works depends on how the system translates search queries to match documents, but generally prompt engineering style “searches” work best with Chatbots that also search rather than straight out search engines

Search engines with LLM vs LLMs that search

Elicit

Add information about all papers

Abstract summary ✓
Intervention
Outcomes measured
Number of participants

Search for paper information

	Paper title
★	The Poetry of Prime Numbers Sarah Glaz 2011 3 Citations
★	On the Infinity of Primes of the Form $2x^2-1$ Pingyuan Zhou 2014 0 Citations

assistant by scite Beta

Install extension!

Title: The Infinite Dance of Primes

Abstract:

This poem explores the concept of the infinity of primes, delving into the mysterious and captivating nature of these elusive numbers. It takes the reader on a journey through the vast expanse of the mathematical realm, where primes dance in an eternal symphony. Through vivid imagery and rhythmic language, the poem seeks to evoke a sense of wonder and awe at the infinite beauty of prime numbers.

Introduction:

In the realm of numbers, where patterns unfold,
There lies a mystery, a tale yet untold.
Primes, those enigmatic beings divine,
Infinite in number, forever they shine.

Verse 1:

Like stars in the night, they twinkle and gleam,
Primes, the building blocks of every number scheme.
From two to three, and five, seven, and beyond,
Their infinite dance, forever goes on.

Verse 2:

No formula can predict their sacred flight,
Primes, elusive and hidden from sight.
They defy all attempts to be tamed,
Their secrets, forever, remain unnamed.

Verse 3:

In the vast expanse of the number line,
Primes emerge, like gems that brightly shine.
They stand alone, indivisible and pure,
Their beauty eternal forever endure

Input : Write a poem on infinity of primes (Elicit vs Scite assistant)

2. Generation of direct answers

Generation of answers - ChatGPT alone can't cut it



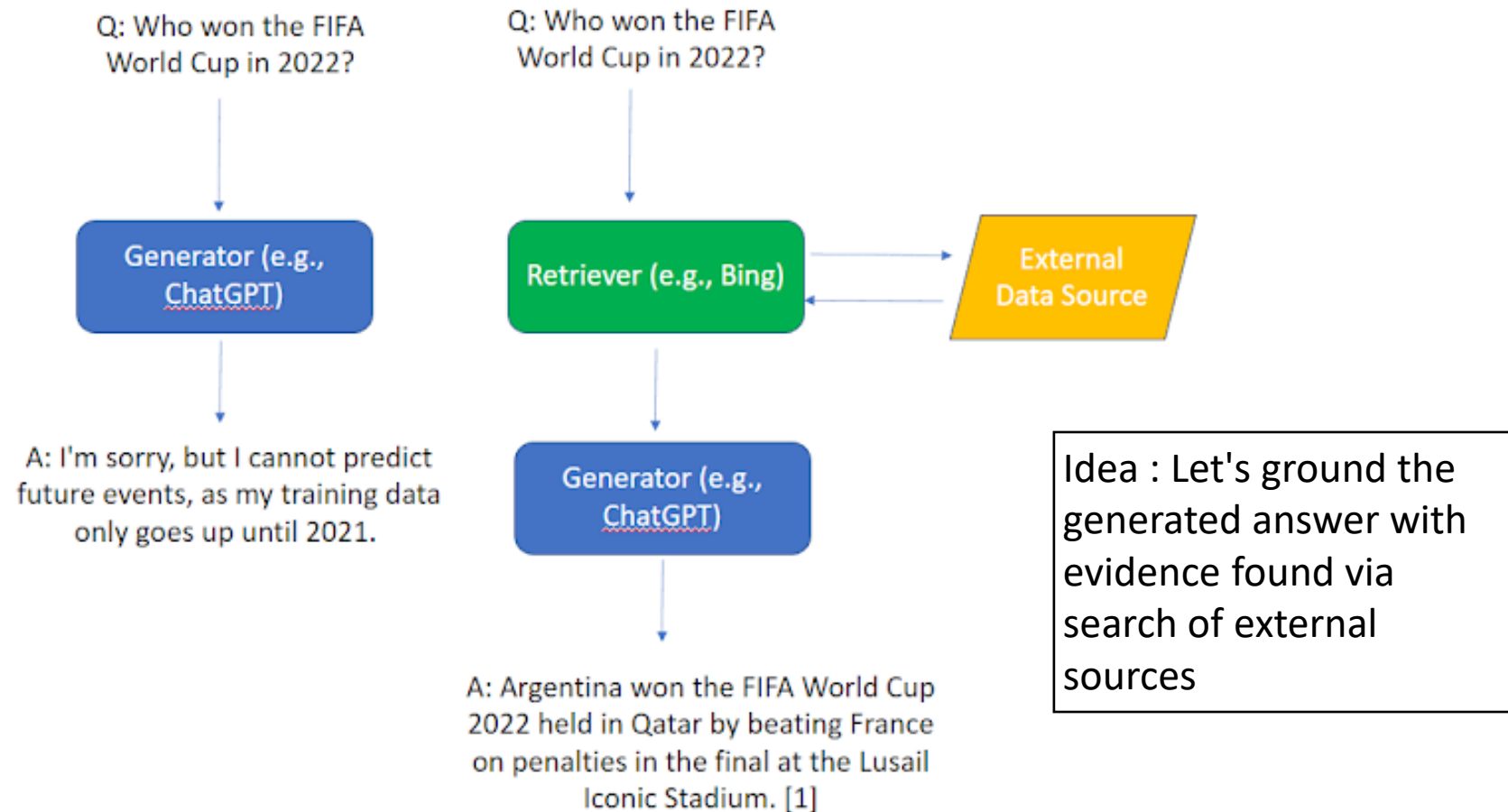
who won the fifa world cup in 2022?



I'm sorry, but as of my last training data in September 2021, I don't have the ability to provide information about events that occurred after that time, including the winner of the 2022 FIFA World Cup. To obtain the most accurate and up-to-date information, I recommend checking the latest news sources or the official FIFA website.

1. Lack of currency in data trained on
2. Even with low hallucinations there is no way to verify answers

Retriever augmented generation (RAG) to rescue



Possible basic solution



One academic example with details

CORE+GPT

CORE-GPT: Combining Open
Access research and AI for
credible, trustworthy question
answering

Is the COVID vaccine safe?

According to various studies and reports, COVID-19 vaccines are considered safe and effective in preventing COVID-19 infection. Pregnant and lactating women have shown positive results of the COVID-19 vaccine, with immune responses detected in maternal blood, cord blood, and breast milk [url:<https://core.ac.uk/works/126745019>]. In a conjoint experiment conducted in France, Germany, and Sweden, vaccine preferences among people are mainly based on differences in effectiveness and side-effects rather than the country of origin or vaccine technology. [url:<https://core.ac.uk/works/124504363>] A cross-sectional survey, monitoring short-term side effects among specific population groups of COVID-19 vaccines, has been globally administered. The survey is set to monitor the long-term safety and effectiveness of the vaccine over multiple phases [url:<https://core.ac.uk/works/121751576>]. Reports have suggested that through clinical trials, the effectiveness and safety of the COVID-19 vaccine has been shown, reducing the risk of COVID-19 infection, and promoting public health [url:<https://core.ac.uk/works/122628784>].

1. <https://core.ac.uk/works/18613179> - Building confidence in COVID-19 vaccines among your patients : tips for the healthcare team
2. <https://core.ac.uk/works/124504363> - Which vaccine attributes foster vaccine uptake? A cross-country conjoint experiment
3. <https://core.ac.uk/works/126745019> - EFFICACY, IMMUNOGENICITY AND SIDE EFFECT OF COVID-19 VACCINE ON PREGNANT AND LACTATING WOMEN: A SYSTEMATIC REVIEW
4. <https://core.ac.uk/works/122628784> - Efektivitas Dan Keamanan Vaksin Covid-19 : Studi Refrensi
5. <https://core.ac.uk/works/121751576> - COVID-19 Vaccines Safety Tracking (CoVaST): Protocol of a Multi-Center Prospective Cohort Study for Active Surveillance of COVID-19 Vaccines' Side Effects.

[See more in CORE](#)

CORE GPT - <https://blog.core.ac.uk/2023/03/17/core-gpt-combining-open-access-research-and-ai-for-credible-trustworthy-question-answering/>

Search + LLM (CoreGPT)

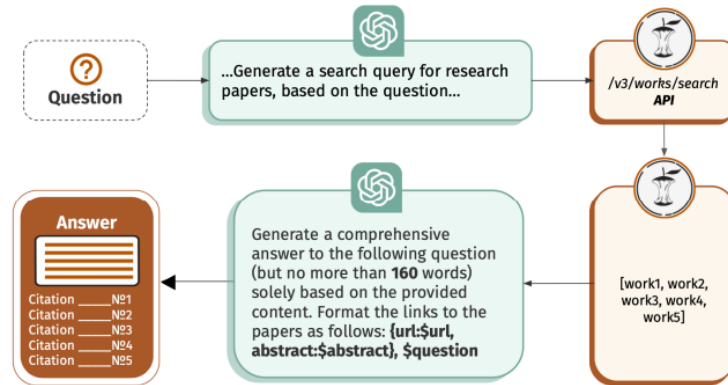


Fig. 2. CORE-GPT workflow.

- Identify the key terms within the question
- Enrich with close synonyms
- Formulate this into a search query.

A sample question and search formatted response can be seen below:

Original user question

What strategies can be implemented to improve literacy rates in rural primary schools in developing countries?

Formatted query

strategies improve literacy rates rural primary schools developing countries OR low-income OR underdeveloped OR third-world

In Stage 2, the formatted search query is then passed to the CORE API which returns the five most relevant papers where the full-text content is available. Stage 3 is the key to the novel solution provided by CORE-GPT. We pass the titles and abstracts returned in Stage 2 back to the GPT4 API with further instructions:

Generate a comprehensive answer to the following question (but no more than 160 words) solely based on the content provided. Format the links to the papers as follows: furl:Surl, abstract:\$abstract, \$question

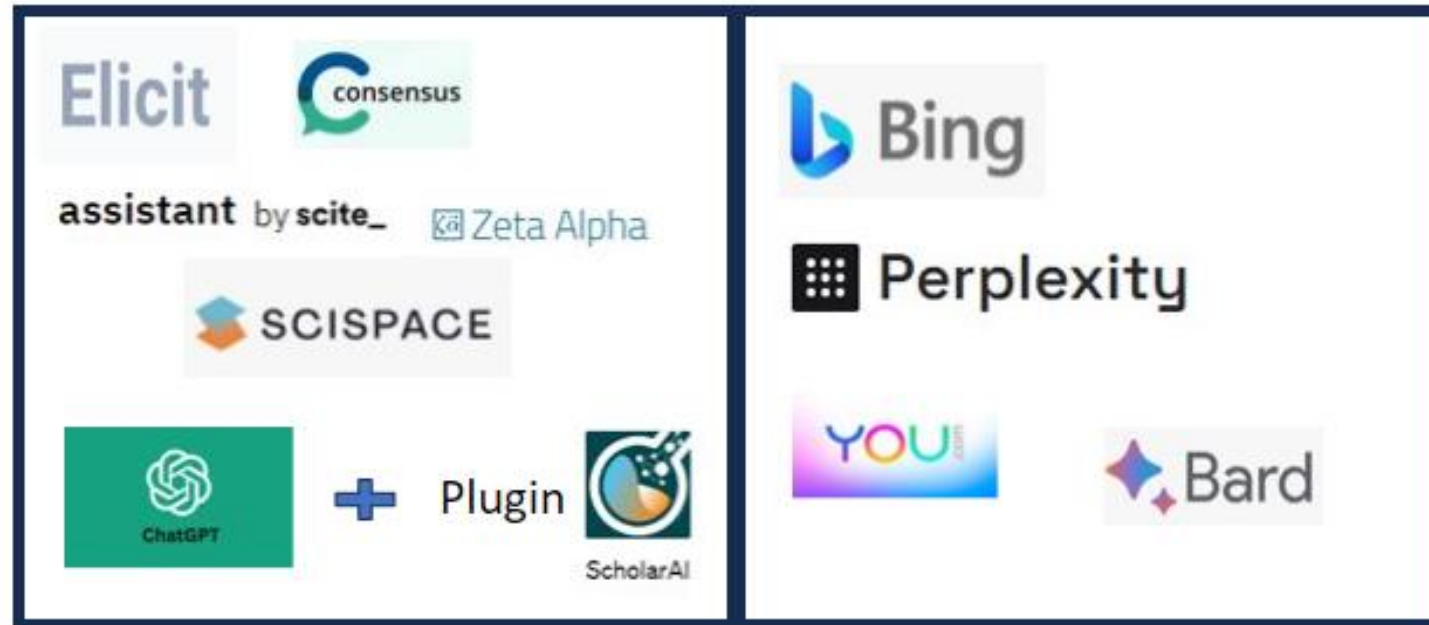
<https://arxiv.org/abs/2307.04683>

Academic vs General Web

Retriever Augmented Language Models

Academic

General Web



More are coming....

ChatGPT-like AIs are coming to major science search engines

The Scopus, Dimensions and Web of Science databases are introducing conversational AI search.

[Richard Van Noorden](#)



Learn with AI-generated overviews based on documents since 2018 [How it works](#)

1 Influence of seismology on civil engineering designs

2 Influence of seismology on civil engineering designs

Seismology plays a crucial role in civil engineering designs. It helps in understanding the behavior of engineering structures under earthquake effects and determining the location of seismic stations [1]. Seismic waves' velocity is important for defining suitable construction locations and monitoring seismic activity efficiently [2]. Seismology also contributes to the safety, sustainability, and resilience of civil engineering structures in seismic areas [3]. It aids in the innovation and design of exceptional structures in earthquake-prone areas [4]. Additionally, seismological research provides valuable insights into crustal architecture, geodynamics, and earthquake source parameters, benefiting engineering seismology applications [5]. Overall, seismology significantly influences civil engineering designs by enhancing safety, efficiency, and resilience in earthquake-prone regions [2] [3] [6].

Show all references

3 How does seismology influence the design of tall buildings in earthquake-prone areas?

4 What role does seismology play in the design of bridges to ensure their stability during seismic events?

5 How does seismology impact the construction of underground structures like tunnels and subway systems?

Share feedback

4

Earthquake monitoring

Geotechnical Engineering

Earthquakes

Structural analysis

Civil Engineering

Seismic Design

Damage Study

Impact

Research

Sustainability

Innovation

Dimensions

Dimensions AI Assistant (beta)

is covid-19 airborne?

Summary

The evidence suggests that covid-19 may be spread through airborne transmission, but further research is needed to confirm this [1]. The current review suggests that there is a potential association between pollen bioaerosols and COVID-19, but more research is needed to show how pollen bioaerosols could affect virus survivals [2]. The text suggests that the potential of airborne transmission of COVID-19 under eased restrictions has not been considered important enough, suggesting that it is airborne [3]. The evidence suggests that COVID-19 virus can potentially spread by airborne transmission, and the results of the mechanical model encourage further research on the deliberate moisturization of room air as a potential technique for control of airborne COVID-19 transmission [4].

<https://www.nature.com/articles/d41586-023-02470-3>

<https://beta.elsevier.com/products/scopus/scopus-ai>

One-example - Elicit.org

Elicit

is there an open access citation advantage?

SUMMARY OF TOP 4 PAPERS **BETA**

The papers collectively suggest that there is a citation advantage for open access articles, but the extent of the advantage varies by discipline. [Norris 2008](#) found that open access articles had a citation advantage over toll access articles in four disciplines, with sociology having the highest citation advantage. [Craig 2007](#) reviewed the literature and found that the citation advantage of open access articles is likely due to selection bias rather than the open access status alone. [Swan 2010](#) provided a summary of reported studies on the open access citation

	Paper title
	<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>The Open Access citation advantage: Studies and results to date</p> <p> A. Swan</p> <p>2010 214 Citations</p>
	<p>Do open access articles have greater citation impact?: A critical r</p> <p> Iain D. Craig, Andrew M. Plume, Marie E. McVeigh, James Prin</p> <p> <i>I. Informatics</i></p>

ChatGPT + Scholar.io plugin



Used ScholarAI ^

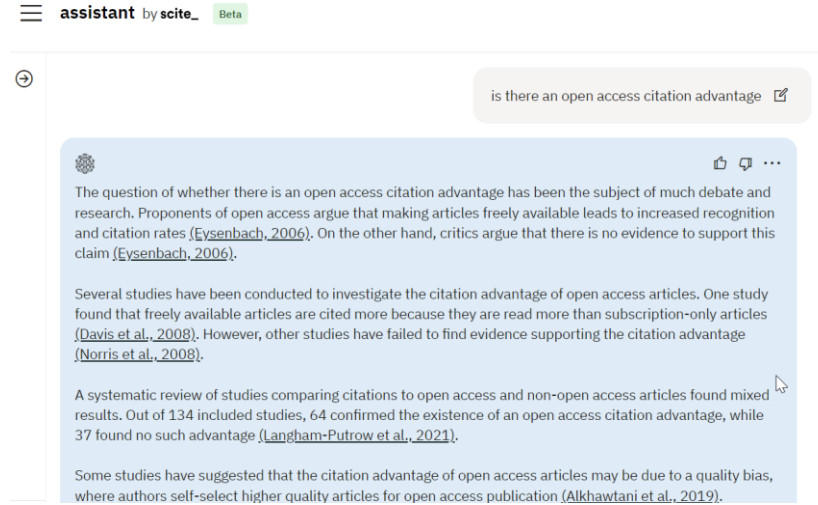
REQUEST TO SCHOLARAI

```
{
  "keywords": "open access, citation advantage",
  "query": "open access citation advantage"
}
```

RESPONSE FROM SCHOLARAI

```
{
  "next_offset": 4,
  "paper_data": [
    {
      "abstract": "Aims Over the last two decades, the existence of an open access citation advantage (OACA)—increased citation of articles made available open access (OA)—has been the topic of much discussion. While there has been substantial research to address this question, findings have been contradictory and inconclusive. We conducted a systematic review to compare studies of citations to OA and non-OA articles. Methods A systematic search of 17 databases attempted to capture all relevant studies authored since 2001. The protocol was registered in Open Science Framework. We included studies with a direct comparison between OA and non-OA items and reported article-level citation as an outcome. Both randomized and non-randomized studies were included. No limitations were placed on study design, language, or publication type. Results A total of 5,744 items were retrieved. Ultimately, 134 items were identified for inclusion. 64 studies (47.8%) confirmed the existence of OACA, while 37 (27.6%) found that it did not exist, 32 (23.9%) found OACA only in subsets of their sample, and 1 study (0.8%) was inconclusive. Studies with a focus on multiple
```

Others



is there an open access citation advantage

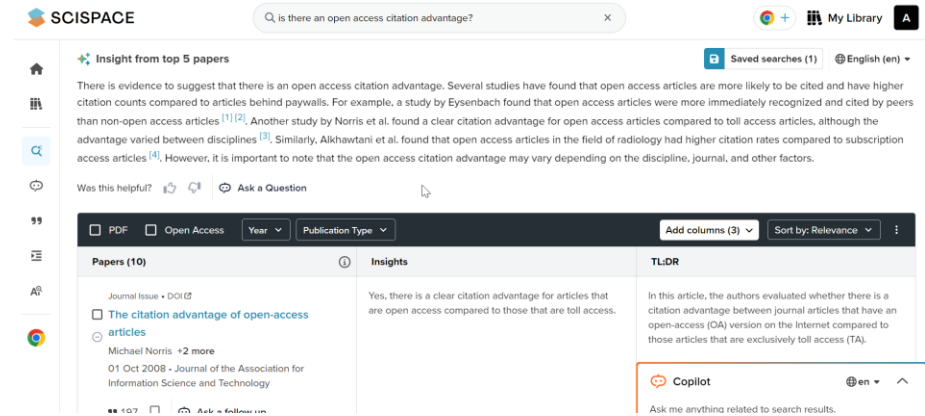
The question of whether there is an open access citation advantage has been the subject of much debate and research. Proponents of open access argue that making articles freely available leads to increased recognition and citation rates (Eysenbach, 2006). On the other hand, critics argue that there is no evidence to support this claim (Eysenbach, 2006).

Several studies have been conducted to investigate the citation advantage of open access articles. One study found that freely available articles are cited more because they are read more than subscription-only articles (Davis et al., 2008). However, other studies have failed to find evidence supporting the citation advantage (Norris et al., 2008).

A systematic review of studies comparing citations to open access and non-open access articles found mixed results. Out of 134 included studies, 64 confirmed the existence of an open access citation advantage, while 37 found no such advantage (Langham-Putrow et al., 2021).

Some studies have suggested that the citation advantage of open access articles may be due to a quality bias, where authors self-select higher quality articles for open access publication (Alkhawtani et al., 2019).

Scite assistant



SCISPAC

is there an open access citation advantage?

Insight from top 5 papers

There is evidence to suggest that there is an open access citation advantage. Several studies have found that open access articles are more likely to be cited and have higher citation counts compared to articles behind paywalls. For example, a study by Eysenbach found that open access articles were more immediately recognized and cited by peers than non-open access articles [1][2]. Another study by Norris et al. found a clear citation advantage for open access articles compared to toll access articles, although the advantage varied between disciplines [3]. Similarly, Alkhawtani et al. found that open access articles in the field of radiology had higher citation rates compared to subscription access articles [4]. However, it is important to note that the open access citation advantage may vary depending on the discipline, journal, and other factors.

Was this helpful? Ask a Question

Papers (10)

Journal Issue • DOI 10.1108/JASIS-01-2008-0001

The citation advantage of open-access articles

Michael Norris +2 more

01 Oct 2008 - Journal of the Association for Information Science and Technology

Insights

Yes, there is a clear citation advantage for articles that are open access compared to those that are toll access.

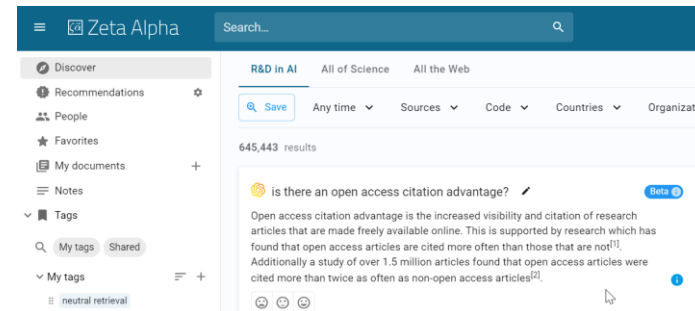
TLDR

In this article, the authors evaluated whether there is a citation advantage between journal articles that have an open-access (OA) version on the Internet compared to those articles that are exclusively toll access (TA).

Copilot

Ask me anything related to search results.

scispace



Zeta Alpha

Search...

Discover

Recommendations

People

Favorites

My documents

Notes

Tags

My tags

Shared

My tags

neutral retrieval

R&D in AI

All of Science

All the Web

Save

Any time

Sources

Code

Countries

Organization

645,443 results

is there an open access citation advantage?

Open access citation advantage is the increased visibility and citation of research articles that are made freely available online. This is supported by research which has found that open access articles are cited more often than those that are not[1]. Additionally a study of over 1.5 million articles found that open access articles were cited more than twice as often as non-open access articles[2].

Zeta-alpha

Using Large language models to generate and extract direct answers - More academic search systems - Scite Assistant , Scispace, Zeta Alpha

Besides academic search, they work as chat support

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. corp.gov.law.harvard.edu
2. investopedia.com
3. bdo.com
+2 more

2 of 8 ●

It can look into open access papers on our institutional repository


what are some datasets for ceo compensation site:smu.edu.sg

PERPLEXITY

Some datasets for CEO compensation can be found in the following sources:

1. Execucomp database, which provides CEO characteristics and compensation data ¹.
2. Compustat segment database, which offers financial and market data for companies, including executive compensation information ¹.

These datasets are often used in academic research to study the relationship between CEO compensation and various factors, such as stock-based compensation ², globalization ³, and network ties between CEOs and board members ⁴.

1  smu

2  smu

3  smu

4  smu

Sources

 smu

1. Generalist CEOs and audit pricing – SMU InK

Compustat segment database, CEO characteristics and compensation data from the Execucomp database, stock return data from the Center for Research in ...

 smu

2. The effect of CEO stock-based compensation on pricing of ...

We find that a CEO's stock-based compensation strengthens the association ... compensation trends using an extensive dataset of large firms.

 smu

3. Globalization and top income shares – SMU InK

With this new dataset, we document that globalization disproportionately benefits the top executives relative to the workers within the same ...

 smu

4. The downside of the network ties between CEO/CFOs and ...

Hwang and Kim (2009) document that CEOs with socially connected boards receive higher compensation and exhibit weaker pay-performance and turnover-performance ...

Limitations of these tools

1. Very new, accuracy, reliability is unclear
2. Current tools have limited data
(most of these tools currently rely heavily on open scholarly metadata and open access papers)

What does the research say?

Paper talking about how adding augmented retriever helps

Table 1: Examples of Four Types of Errors (text in brackets represents the groundtruth, text in blue indicates the entities compared and their information, text in red indicates where the model (GPT-3.5) makes mistakes).

	# Correct	# Wrong	Comprehension	Factualness	Specificity	Inference
GPT-3.5	121	79	16	31	14	18
GPT-3.5 w/ Evi	185	15	3	0	1	11
GPT-4	138	62	8	28	9	14

Table 2: Number of Types of Failures under Different Settings

When correct evidence is manually added to GPT-3.5, it outperforms not just GPT3-5 but also GPT-4. It becomes 85% accurate more than GPT3-5 (60.5%) alone and GPT4 (72%)

03

What does the research say?

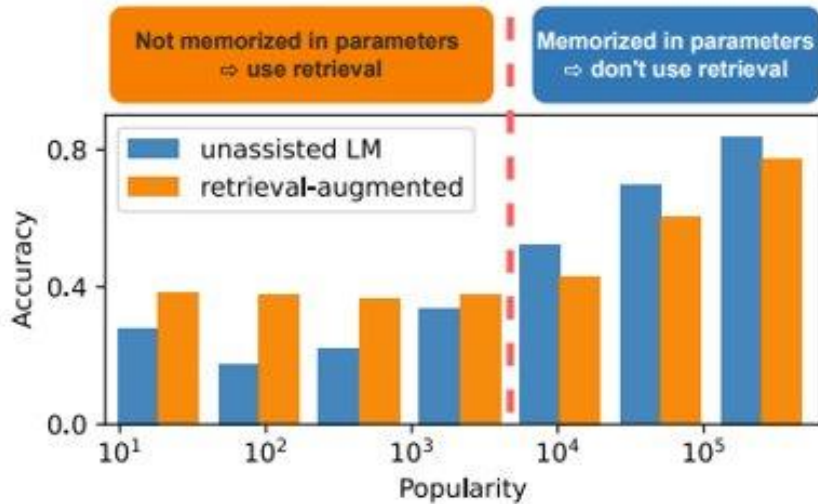


Figure 1: Relationship between subject entity popularity in a question and GPT-3 davinci-003 performance in entity-centric open-domain question answering, with and without retrieved passages from BM25.

Question (answer)	LM	Retrieval-LM	Retrieval
Who was the director of Black? (Sanjay Leela Bhansali)	The director of Black is Sanjay Leela Bhansali. ✓	Noel Black ✗	Noel Black (June 30, 1937 – July 5, 2014) was an American film and television director, screenwriter, and producer. Black was born in Chicago, Illinois. He won awards at the 1966 ... and subsequently concentrated on directing for television, occasionally directing films such as "Private School".
Who was the producer of The Faculty? (Robert Rodriguez)	The Faculty was produced by Elizabeth Avellan and Robert Rodriguez. ✓	The producer of The Faculty was Elizabeth Avellan. ✗	The Faculty is a 1998 American science fiction teen horror film written by Kevin Williamson, directed by <u>Robert Rodriguez</u> , and starring Elijah Wood,... The film was released on Christmas Day, 1998, by Miramax Films through Dimension Films. ... The film has developed a cult following in the years since its release.
What is Michael Shelley's occupation? (singer-songwriter)	Michael Shelley is a singer-songwriter and musician. ✓	Michael Shelley is an American applied mathematician. ✗	Michael Shelley (mathematician) Michael J. Shelley (born August 17, 1959) is an American applied mathematician who works on the modeling and simulation of complex systems arising in physics and biology. This has included free-boundary problems in fluids and materials science... He is also the co-founder and co-director of the Courant Institute's Applied Mathematics Lab.
In what city was Zijah Sokolović born? (Sara-jevo)	Zijah Sokolović was born in Sarajevo. ✓	Zijah Sokolović was born in Orahovac, Kingdom ✗	Ali Sokol ... (born 8 May 1921 in Orahovac, Kingdom of Serbs, Croats and Slovenes, died 23 September 1974) was a Yugoslav pulmonologist . Ali Sokol was born into an agricultural family. He was the third of four children of father and mother Hatixhes Solomon. It is unknown the exact date of birth but the most reliable date is May 8 year in 1921.

Table 3: Qualitative examples of the questions where only retrieval-augmented LMs *fail* to answer correctly. The blue underlined text indicates the sub-strings matching the gold answers in the retrieved context.

What does the research say?

Paper comparing Bing Chat, Perplexity, Youchat and NeevaAI

51.5% of generated sentences are fully supported by citations and only 74.5% of citations support their associated sentence. (Average of 4 search engines)

68.7% of generated sentences are fully supported by citations and only 89.5% of citations support their associated sentence. (Best of 4 search engines)

First generated statement [1✓][2✗][3⚠].
Second generated statement [1✓][2✗][4✗].
Third generated statement [4✓][5⚠].

Citation Recall: 3/3 = 100%
Citation Precision: 3/8 = 37.5%

First generated statement [1⚠][2⚠].
Second generated statement [2✗].
Third generated statement.

Citation Recall: 1/3 = 33%
Citation Precision: 2/3 = 66%

First generated statement [1✓][2✓][3✗].
Second generated statement.
Third generated statement.

Citation Recall: 1/3 = 33%
Citation Precision: 2/3 = 66%

✓: highlighted statement is fully supported by citations
✗: highlighted statement is not fully supported by citations.
✓: citation fully supports its associated statement.
⚠: citation partially supports its associated statement.
✗: citation does not support its associated statement.

<i>Citation Precision (%; ↑)</i>		<i>Citation Precision (%; ↑)</i>					
Average Over All Queries		AllSouls	davinci-debate	ELI5		WikiHowKeywords	
				KILT	Live		
Bing Chat	89.5	Bing Chat	88.8	88.8	87.6	87.2	92.1
NeevaAI	72.0	NeevaAI	69.8	74.1	75.7	73.8	74.0
perplexity.ai	72.7	perplexity.ai	61.7	68.4	64.9	66.3	77.4
YouChat	63.6	YouChat	51.1	50.0	64.7	57.9	71.1
Average	74.5	Average	67.8	70.3	73.2	71.3	78.7

<i>Citation Precision (%; ↑)</i>							
NaturalQuestions							
	List Long Answer		Table Long Answer		Paragraph Long Answer		No Answer
	Has Short	No Short	Has Short	No Short	Has Short	No Short	
Bing Chat	86.8	86.8	89.0	92.5	92.9	91.3	90.8
NeevaAI	73.2	67.6	67.1	64.2	73.4	76.5	70.8
perplexity.ai	82.1	81.0	76.0	71.7	83.8	79.7	74.0
YouChat	63.3	62.7	64.8	56.1	75.7	67.5	58.6
Average	76.4	74.5	74.2	71.1	81.5	78.7	73.5

Table 6: Human evaluation results of citation precision in existing generative search engines. Citation precision is concerningly low (many generated citations do not support their associated statements), especially given that these systems already have millions of users and may serve as a primary tool for fulfilling user information needs.

<i>Citation Recall</i> (%; ↑)		<i>Citation Recall</i> (%; ↑)					
Average Over All Queries		AllSouls	davinci-debate	ELI5		WikiHowKeywords	
				KILT	Live		
Bing Chat	58.7	Bing Chat	55.6	57.1	59.8	59.9	50.7
NeevaAI	67.6	NeevaAI	55.3	66.3	66.6	61.6	72.5
perplexity.ai	68.7	perplexity.ai	63.0	64.2	64.8	58.1	74.6
YouChat	11.1	YouChat	3.2	3.9	3.0	4.6	12.1
Average	51.5	Average	44.3	47.9	48.5	46.0	52.5

<i>Citation Recall</i> (%; ↑)							
NaturalQuestions							
	List Long Answer		Table Long Answer		Paragraph Long Answer		No Answer
	Has Short	No Short	Has Short	No Short	Has Short	No Short	
Bing Chat	74.1	60.6	63.5	49.2	72.1	66.3	61.9
NeevaAI	73.0	64.2	69.5	65.1	75.0	74.8	65.6
perplexity.ai	85.3	74.4	79.6	62.4	84.9	75.9	68.4
YouChat	21.6	16.6	30.6	11.5	31.6	21.8	17.8
Average	63.5	53.9	60.8	47.1	65.9	59.7	53.4

Table 5: Human evaluation results of citation recall in existing generative search engines. Citation recall is concerningly low (many generated statements are not fully supported by citations), especially given that these systems already have millions of users and may serve as a primary tool for fulfilling user information needs.

Will users overly trust generated answers?

<i>Pearson Correlation (r)</i>		
	Fluency	Perceived Utility
Citation Recall	-0.76	-0.53
Citation Precision	-0.84	-0.96

Table 8: Citation recall and precision are inversely correlated with fluency and perceived utility in existing generative search engines—responses that are fluent and appear more helpful often have *lower* citation recall and precision.

Liu, N. F., Zhang, T., & Liang, P. (2023). *Evaluating verifiability in generative search engines*. arXiv.
<http://arxiv.org/abs/2304.09848>

Why there is a negative correlation?

Fluency – The more the generator (chatGPT) tries to paraphrase the more fluent the answer is, but likely to lead to the generated sentence not being supported by citation



is covid-19 airborne?

Aerosol and Air Quality Research | Mostafa Hadei et al. | 2020

Save Cite Share

Consensus used AI to add missing context from the paper to this finding.

The study found that SARS-CoV-2 RNA was not detectable by air samplers, suggesting that the airborne route is not the predominant mode of transmission of COVID-19, but this conclusion may not apply during aerosol-generating procedures or in cohort wards with large numbers of COVID-19 patients.

Air and environmental sampling for SARS-CoV-2 around hospitalized patients with coronavirus disease 2019 (COVID-19)

Infection Control and Hospital Epidemiology | V. Cheng et al. | 2020

Highly Cited

Save Cite Share

The presence of the virus in the air beyond a 1-m distance from the patients and surfaces of the hospital indicates that the SARS-CoV-2 virus has the potential to be transmitted by airborne and surface routes from COVID-19 patients to health-care workers working in COVID-19 dedicated hospital.

Evidence of the presence of SARS-CoV-2 virus in atmospheric air and surfaces of a dedicated COVID hospital

Journal of Medical Virology | A. Dubey et al. | 2021

Save Cite Share


Sentence is paraphrased

Sentence is exactly extracted from paper

<https://consensus.app/results/?q=is%20covid-19%20airborne%3F>

Why there is a negative correlation

Perceived usefulness – Assume no good answer is found. The more the generator (chatgpt) tries to generate sentences supported by evidence found , the less likely the generated sentence will answer the question and hence lead to lower perceived usefulness.

Write an essay on how board composition of banks influences risk-taking 



The composition of a bank's board of directors plays a crucial role in influencing its risk-taking behavior. Research has shown that the diversity and expertise of board members can significantly impact a bank's risk management practices and decision-making processes (Zhang, 2012). A board with a diverse range of skills, knowledge, and experience can provide valuable insights and perspectives that help identify and mitigate potential risks (Ackerman et al., 2011). Additionally, studies have found that boards with independent directors who possess relevant industry expertise are more effective (Ting & Yang, 2022). Furthermore, the presence of directors with financial expertise can enhance decision-making (Reb et al., 2016). Overall, a well-composed board that includes individual expertise within a bank's management framework can lead to more informed risk-related decisions and a more robust risk management framework.

Not what you were looking for? [Change Assistant Settings](#)

"...Let's get serious: Communicating commitment in romantic relationships...."

Let's get serious: Communicating commitment in romantic relationships.

Ackerman¹, Griskevicius², Li³ 2011
Journal of Personality and Social Psychology

 65 |  4 |  39 |  0

[View full text](#) [Add to dashboard](#) [Cite](#)

**3. Extraction of information
from papers (abstract and full-
text) to enhance search engine
result pages
(Also "Chat with PDF")**

An example of extracting data -> elicit.org (beta version out soon)

Elicit

FAQ

Tasks

is there an open access citation advantage?

SUMMARY OF TOP 4 PAPERS

BETA

The papers suggest that there is a citation advantage for open access articles, but the size of the advantage varies by discipline and the quality of the article. [Ottaviani 2016](#) found that open access articles have a citation advantage of up to 19%, while [Norris 2008](#) found that open access articles have a clear citation advantage over toll access articles, but the advantage varies by discipline.

Add information about all papers

Outcomes measured ✓

Number of participants

Detailed study design ✓

Search for paper information

	Paper title	Outcomes measured	Detailed study design
☆	<div>The Post-Embargo Open Access Citation Advantage: It Exists (Probably), It's Modest (Usually), and the Rich Get Richer (of Course)</div> <div> <div>J. Ottaviani</div> <div> <div>PLoS ONE</div> <div></div> </div> </div> <div> <div>2016</div> <div>63 Citations</div> <div>DOI</div> <div>PDF</div> </div>	<ul style="list-style-type: none"> •Number Of Downloads Of Open Access Articles •Number Of Citations Of Open Access Articles 	-
☆	<div>The citation advantage of open-access articles</div> <div> <div>Michael Norris, C. Oppenheim, Fytton Rowland</div> <div> <div>J. Assoc. Inf. Sci. Technol.</div> </div> </div> <div> <div>2008</div> <div>502 Citations</div> <div>DOI</div> <div>PDF</div> </div>	<ul style="list-style-type: none"> •Citation Count •Oa Status 	observational
☆	<div>The citation advantage of open-access articles</div> <div> <div>Michael Norris, Charles Oppenheim, Fytton Rowland</div> </div> <div> <div>2008</div> <div>124 Citations</div> <div>DOI</div> </div>	<ul style="list-style-type: none"> •Citation Count •Oa Status 	observational
☆	<div>Revisiting the Open Access Citation Advantage for Legal Scholarship</div> <div> <div>John R. Beatty</div> </div>	<ul style="list-style-type: none"> •The Citation Advantage For Open Access Legal Scholarship 	-

You are not just limited to predefined criteria/categories

Elicit

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

is there an open access citation advantage?

SUMMARY OF TOP 4 PAPERS

BETA

The papers suggest that there is a citation advantage for open access articles, but the size of the advantage varies by discipline and the quality of the article. [Ottaviani 2016](#) found that open access articles have a citation advantage of up to 19%, while [Norris 2008](#) found that open access articles have a clear citation advantage over toll access articles, but the advantage varies by discipline. [Beatty 2019](#) notes that the definition of open access and the source of citation data can affect the results of open access citation studies. Overall, the papers suggest that there is an open access citation advantage, but the size of the advantage and the factors that influence it are

Add information about all papers

Intervention

Outcomes measured ✓

Number of participants

Detailed study design ✓

What was the discipline covered? ✓

Search for paper information

Has PDF

Filter

Sort by

Export as

Paper title	Abstract summary	Outcomes measured	Detailed study design	What was the discipline covered?
<div>☆</div> <div>The Post-Embargo Open Access Citation Advantage: It Exists (Probably), It's Modest (Usually), and the Rich Get Richer (of Course)</div> <div> <div>J. Ottaviani</div> <div>PLoS ONE</div> </div> <div>2016 63 Citations</div> <div>DOI</div> <div>PDF</div>	Open access articles are cited more often than closed access/subscription-only articles.	<ul style="list-style-type: none"> Number Of Downloads Of Open Access Articles Number Of Citations Of Open Access Articles 	-	-
<div>☆</div> <div>The citation advantage of open-access articles</div> <div> <div>Michael Norris, C. Oppenheim, Fytton Rowland</div> <div>J. Assoc. Inf. Sci. Technol.</div> </div> <div>2008 502 Citations</div> <div>DOI</div> <div>PDF</div>	There appears to be a clear citation advantage for those articles that are OA as opposed to those that are TA.	<ul style="list-style-type: none"> Citation Count Oa Status 	observational	ecology, applied mathematics, sociology, and economics
<div>☆</div> <div>The citation advantage of open-access articles</div> <div> <div>Michael Norris, Charles Oppenheim, Fytton Rowland</div> </div> <div>2008 124 Citations</div> <div>DOI</div>	There appears to be a clear citation advantage for those articles that are OA as opposed to those that are TA.	<ul style="list-style-type: none"> Citation Count Oa Status 	observational	ecology, applied mathematics, sociology, and economics
<div>☆</div> <div>Revisiting the Open Access Citation Advantage for Legal Scholarship</div> <div> <div>John R. Beatty</div> </div> <div>2019 4 Citations</div>	Open access legal scholarship has shown a significant citation advantage in law citation studies.	<ul style="list-style-type: none"> The Citation Advantage For Open Access Legal Scholarship 	-	law
<div>☆</div> <div>Do open access journal articles experience a citation advantage? Results and methodological reflections of an application of multiple measures to an analysis by WoS subject areas</div> <div> <div>Ishael Razon, Iana B. Blanckenhorn, Heidi Proszek</div> </div>	Open access journal articles experience a citation advantage in very few subject areas.	<ul style="list-style-type: none"> Normalised Citation Score (Ncs) Whether An Article Was Cited At All Whether An Article Is Among The Most Frequently Cited Percentile Of Articles Within Its Respective Subject Area (Pptopx) 	bibliometric analysis	-

You should verify the extractions.... and "chat with pdf"

The citation advantage of open-access articles

Michael Norris, C. Oppenheim, Fytton Rowland

J. Assoc. Inf. Sci. Technol.

2008 502 Citations [PDF](#) [Semantic Scholar](#) [DOI](#)

Abstract summary

There appears to be a clear citation advantage for those articles that are OA as opposed to those that are TA.

What did they test?

An intervention of open access (OA) version of journal articles on the Internet

What outcomes did they measure?

- Citation Count
- Oa Status

Who were the participants?

Participants' occupations were

What was the discipline covered?

ecology, applied mathematics, sociology, and economics

Copy  

Can I trust this paper?

- This study was observational

Ask a question about this paper

Ask

Four subjectsecology, applied mathematics, sociology, and economicswere selected to assess whether there is a citation advantage between journal articles that have an open-access (OA) version on the Internet compared to those articles that are exclusively toll access (TA). Citations were counted using the Web of Science, and the OA status of articles was determined by searching OAlster, OpenDOAR, Google, and Google Scholar. Of a sample of 4,633 articles examined, 2,280 (49%) were OA and had a mean citation count of 9.04 whereas the mean for TA articles was 5.76. There appears to be a clear citation advantage for those articles that are OA as opposed to those that are TA. This advantage, however, varies between disciplines, with sociology having the highest citation advantage, but the lowest number of OA articles, from the sample taken, and ecology having the highest individual citation count for OA articles, but the smallest citation advantage. Tests of correlation or association between OA status and a number of variables were generally found to weak or inconsistent. The cause of this citation advantage has not been determined. © 2008 Wiley Periodicals, Inc.

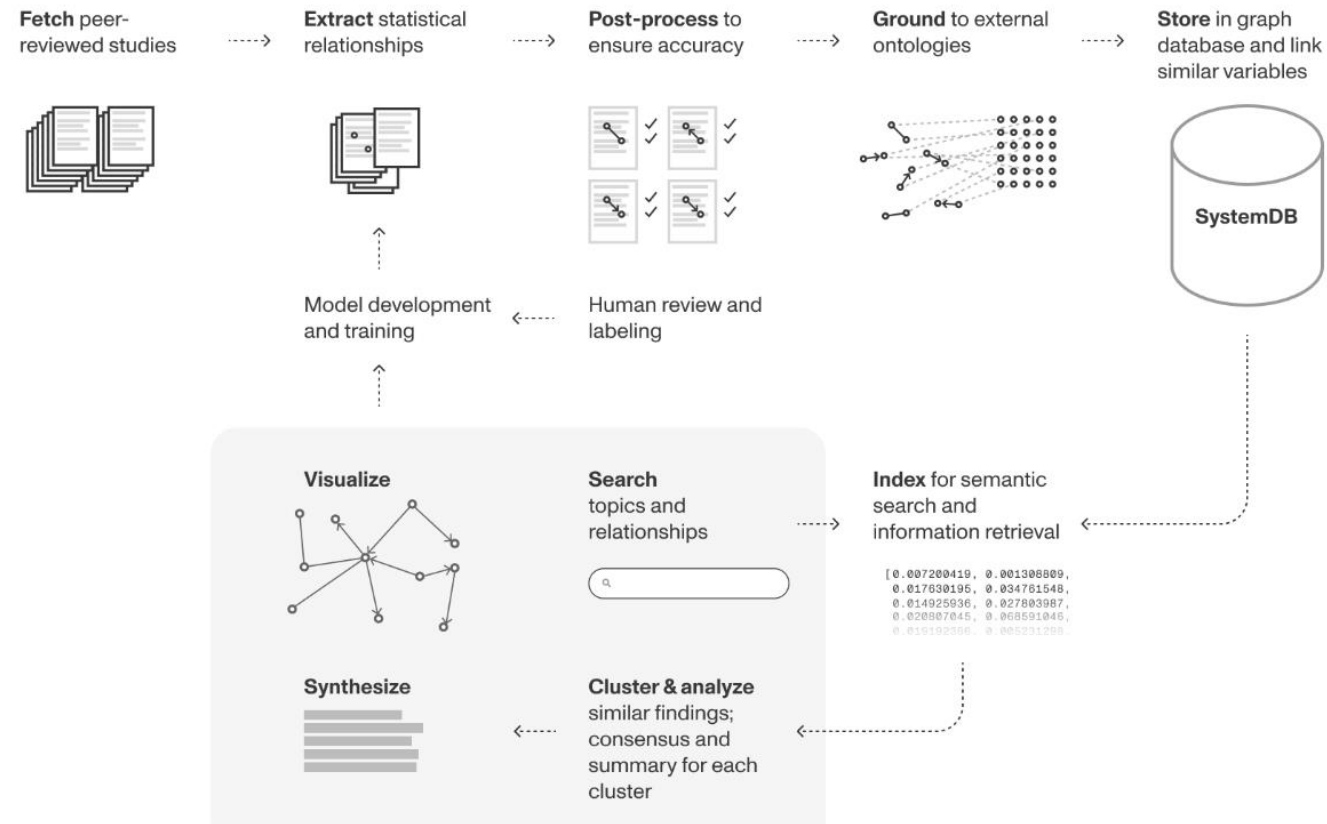
Introduction

Academics are frequently judged, in part at least, on the quality of their published research. The greater the impact of that research as counted by, for example, the number of citations it receives, the better, it is believed, is the quality of the work (van Leeuwen et al. 2003, pp. 262-263) . Receiving many citations for academic research generally correlates strongly with academic success; an analysis of Nobel laureates and their citation counts by Garfield (1979, pp. 63-64) and Opthof (1997, p. 2), although tenuous, gives some credibility to the idea that the two are linked. Likewise a similar ranking by Hirsch (2007, pp. 16569-16572) using his h-index, which uses

Previous research

Open Access Citation Advantage Lawrence (2001) was the first to show that conference articles that were OA and freely available on the World Wide Web were more frequently cited than articles that were offline. Since Lawrence's pioneering work, there have been a number of studies that have demonstrated a similar citation advantage (Antelman, 2004;Eysenbach, 2006;Hajjem et al. 2005;. Harnad and his colleagues (Hajjem et al. 2005;) have carried out large-scale trials where they examined the citation counts of OA and TA articles from the same journals from a database of 14 million articles. In physics and in a range of other subjects, they have found a significant citation advantage for those articles that were OA. In these studies, they identified OA versions of articles either by trawling the web using a computer algorithm or by taking self-archived versions from a disciplinary archive and then compared the citation counts of both OA and TA versions. In contrast to this approach, Antelman (2004) selected four subjects and a relatively small number of articles and manually identified OA versions of articles and their respective citation counts. Again, there was a significant citation advantage for those articles that were OA, but with noticeable variations between subjects.

Another way - System Pro



[https://about.system.com/
blog/reinventing-search-](https://about.system.com/blog/reinventing-search-)

Tools for systematic review - System Pro

Based on findings from 40 peer-reviewed studies ^①

- ① **Endometriosis** is associated with increases in **prematurity** in most available studies. 11 studies ▼
- ② **Endometriosis** is associated with increases in **antepartum hemorrhage, childbirth, placenta previa, pregnancy** in most available studies. 12 studies ▼
- ③ **Endometriosis** is associated with a 44.0% decrease to a 108.0% increase in the odds of **live birth**. 5 studies ^

STUDY TITLE	JOURNAL	AUTHORS	RELEVANT FINDINGS	POPULATION	STUDY TYPE	CITED BY	PUBLISH DATE
Endometriosis, assisted reproduction technology, and risk of adverse pregnancy outcome	Human Reproduction	Stephansson, O., et al	<p>– Endometriosis is associated with no change in the odds of preterm birth after assisted reproduction technology (ART).</p> <p>+1 more</p>	Sex: female Location: sweden Sample size: 1442675	N/A	193	2009
<p>FINDINGS</p> <p>– Endometriosis is associated with no change in the odds of preterm birth after assisted reproduction technology (ART).</p> <p>↗ Endometriosis is associated with a 37.0% increase in the odds of preterm birth without assisted reproduction technology (ART).</p>			<p>STATISTICS</p> <p>Adjusted Odds Ratio 1.240 p-value: N/A 95% CI (0.990 to 1.570)</p> <p>Adjusted Odds Ratio 1.370 p-value: N/A 95% CI (1.250 to 1.500)</p>	<p>PLOT</p>	<p>Suggest revision</p> <p>Suggest revision</p>		
The effect of endometriosis on live birth rate and other reproductive outcomes in ART cycles: a cohort study.	Human Reproduction Open	Muteshi, C., et al	<p>– Mild endometriosis vs. unexplained infertility is associated with no change in the odds of live birth.</p> <p>+3 more</p>	Sex: female	N/A	21	2018
<p>FINDINGS</p> <p>– Mild endometriosis vs. unexplained infertility is associated with no change in the odds of live birth.</p> <p>– Minimal endometriosis vs. unexplained infertility is associated with no change in the odds of live birth.</p> <p>↘ Severe endometriosis vs. unexplained infertility is associated with a 44.0% decrease in the odds of live birth.</p> <p>– Moderate endometriosis vs. unexplained infertility is associated with no change in the odds of live birth.</p>			<p>STATISTICS</p> <p>Odds Ratio 0.650 p-value = 0.070 95% CI (0.400 to 1.040)</p> <p>Odds Ratio 0.890 p-value = 0.580 95% CI (0.600 to 1.330)</p> <p>Odds Ratio 0.560 p-value = 0.009 95% CI (0.350 to 0.870)</p> <p>Odds Ratio 0.870 p-value = 0.500 95% CI (0.570 to 1.310)</p>	<p>PLOT</p>	<p>Suggest revision</p> <p>Suggest revision</p> <p>Suggest revision</p> <p>Suggest revision</p>		

Comparison

	Bing Chat	Perplexity.ai	Elicit.org	Scite assistant	Scispace	Consensus.ai
Source	Bing	Bing	Semantic Scholar	Open metadata + Selected partners	Unclear	Semantic Scholar
Coverage type	General web	General web	Academic journals, Conf paper etc	Academic journals, Conf paper etc	Academic journals, Conf paper etc	Academic journals, Conf paper etc
Coverage (metadata)	unknown	unknown	Roughly 175 million	Similar to Elicit	Unknown but likely to be similar	Roughly 75 million
Coverage (full text)	Bing index-	Bing Index-	Open Access content only	33 million	Open Access content only	Open Access content only
Can summarize papers in structured table (lit review matrix)?	No	No	Yes	No	Yes	Consensus Meter for Yes/No questions on 20 results
Can upload your own PDF?	No	No	Yes	No	No	No
Pros/Special feature	Edge plugin	Perplexity Plugin	Most established, good extraction of data	Allows many customization of sources used	Has extension	Employs extractive summary not just generative

See also "[Five AI Research Tools That Referencing Genuine Sources](#)"

4. Some implications

1. Education for users on use of LLMs



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>

Fake citations, even in papers!



Enrique Orduña Malea
@eomalea



I'm afraid to say that several preprint servers are publishing online papers, which cite publications co-authored by me that do not exist. This is the [@chatgptimpact](#). Google Scholar and ResearchGate are indexing those papers, and their fake citations, by the way.

11:49 PM · Apr 24, 2023 · **21.5K** Views



17. Li, J.; Burnham, J.F.; Lemley, T.; Britton, R.M. Citation analysis: Comparison of Web of Science®, Scopus™, SciFinder®, and Google Scholar. *J. Electron. Resour. Med. Libr.* **2010**, *7*, 196–217.
18. Neff MW, Corley EA. 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology. *Scientometrics* **2009**, *80*, 657–682.
19. Orduna-Malea, E.; Thelwall, M.; Kousha, K.; Delgado López-Cózar, E. Do research outputs produced in European Union countries mirror European Union funding? *Scientometrics* **2017**, *112*, 1857–1871.
20. Huang, M.H.; Huang, W.T.; Chen, D.W. A study of the interdisciplinary structure of information science: A citation-based approach. *J. Inf. Sci.* **2016**, *42*, 162–181.
21. Eom, G.; Yoon, D.; Kim, K.; Jung, J. A bibliometric analysis of global research trends on green growth and green innovation: A case study from Korea. *J. Clean. Prod.* **2020**, *269*, 122169.
22. Fister, I.; Fister, I., Jr.; Yang, X.S. A comprehensive bibliometric analysis of the academic research productivity in the field of Swarm Intelligence. *Eng. Appl. Artif. Intell.* **2018**, *75*, 107–117.

<https://twitter.com/eomalea/status/1650527418577309699>

2. Discoverability in a world of LLMs (without search)

GPTBot

GPTBot is OpenAI's web crawler and can be identified by the following user agent and string.

```
User agent token: GPTBot
Full user-agent string: Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko; compatib
```

Usage

Web pages crawled with the GPTBot user agent may potentially be used to improve future models and are filtered to remove sources that require paywall access, are known to gather personally identifiable information (PII), or have text that violates our policies. Allowing GPTBot to access your site can help AI models become more accurate and improve their general capabilities and safety. Below, we also share how to disallow GPTBot from accessing your site.

Disallowing GPTBot

To disallow GPTBot to access your site you can add the GPTBot to your site's robots.txt:

```
User-agent: GPTBot
Disallow: /
```

Customize GPTBot access

To allow GPTBot to access only parts of your site you can add the GPTBot token to your site's robots.txt like this:

```
1 User-agent: GPTBot
2 Allow: /directory-1/
3 Disallow: /directory-2/
```

Give access to OpenAI's GPTBot crawler

Information needs to be indexed/accessible by the search engine part for RAG systems – Here Bing chat fails because catalog not indexed

I am from SMU Libraries in Singapore, does my library own goal Great leader, dear leader : demystifying North Korea under control the Kim Clan

✓ Searching for: **SMU Libraries Singapore Great leader, dear leader : demystifying North Korea under control the Kim Clan**

✓ Generating answers for you...






I'm sorry, but I couldn't find any information about whether SMU Libraries in Singapore owns the book "Great leader, dear leader: demystifying North Korea under the Kim Clan" by Bertil Lintner. However, it is available in the **National Library (Singapore)**¹. Would you like more information on this?

Learn more: [1. nlb.gov.sg](#) [2. archive.org](#) [3. amazon.sg](#)

1 of 20 ●

Importance of getting your info findable by search engines – Google Bard succeeds

 I am from SMU libraries in Singapore, does my library own Great leader, dear leader : demystifying North Korea under the Kim Clan 

[View other drafts](#) 

Draft 1


Yes, SMU libraries do own the book Great leader, dear leader : demystifying North Korea under the Kim Clan. You can find it here:
https://search.library.smu.edu.sg/discovery/fulldisplay/alma994426502601/65SMU_INST:SMU_NUI


Draft 2

Yes, SMU Libraries has a copy of Great Leader, Dear Leader: Demystifying North Korea Under the Kim Clan by Bertil Lintner. It is available in the following locations:

Draft 3

Yes, the Lee Kong Chian Reference Library at Singapore Management University (SMU) owns a copy of the book "Great Leader, Dear Leader: Demystifying North Korea Under the Kim Clan" by Bertil Lintner. The book is



 Yes, SMU libraries do own the book Great leader, dear leader : demystifying North Korea under the Kim Clan. You can find it here:
https://search.library.smu.edu.sg/discovery/fulldisplay/alma994426502601/65SMU_INST:SMU_NUI

The book is available in the following formats:

- Hardcover
- Ebook
- Audiobook

You can borrow the book or access the ebook or audiobook through the SMU Library website or app.

Reformat data in none-table formats

what is the opening hours of the singapore management university libraries?

✓ Searching for: singapore management university library opening hours

✓ Generating answers for you...

The opening hours of Singapore Management University Libraries are as follows:

- Li Ka Shing Library
 - Monday to Friday: 8:30am to 10:00pm
 - Saturday and Sunday: 10:00am to 6:00pm
 - Public Holidays: Closed
- Kwa Geok Choo Law Library
 - Monday to Friday: 8:30am to 10:00pm
 - Saturday and Sunday: Closed
 - Public Holidays: Closed

I hope this helps! Let me know if you have any other questions.

Learn more: [1. bing.com](#) [2. library.smu.edu.sg](#) [3. library.sim.edu.sg](#) [+3 more](#) 1 of 20

TUESDAY, MARCH 21, 2023	
Li Ka Shing Library	8:00am – 11:59pm
Learning Commons	24/7
Service Desk	8:30am – 7:00pm
Kwa Geok Choo Law Library	8:00am – 11:59pm
Learning Commons	24/7
Service Desk	8:30am – 7:00pm
"Ask Library" Chat	10:00am – 7:00pm

Weekly cleaning @ Learning Commons

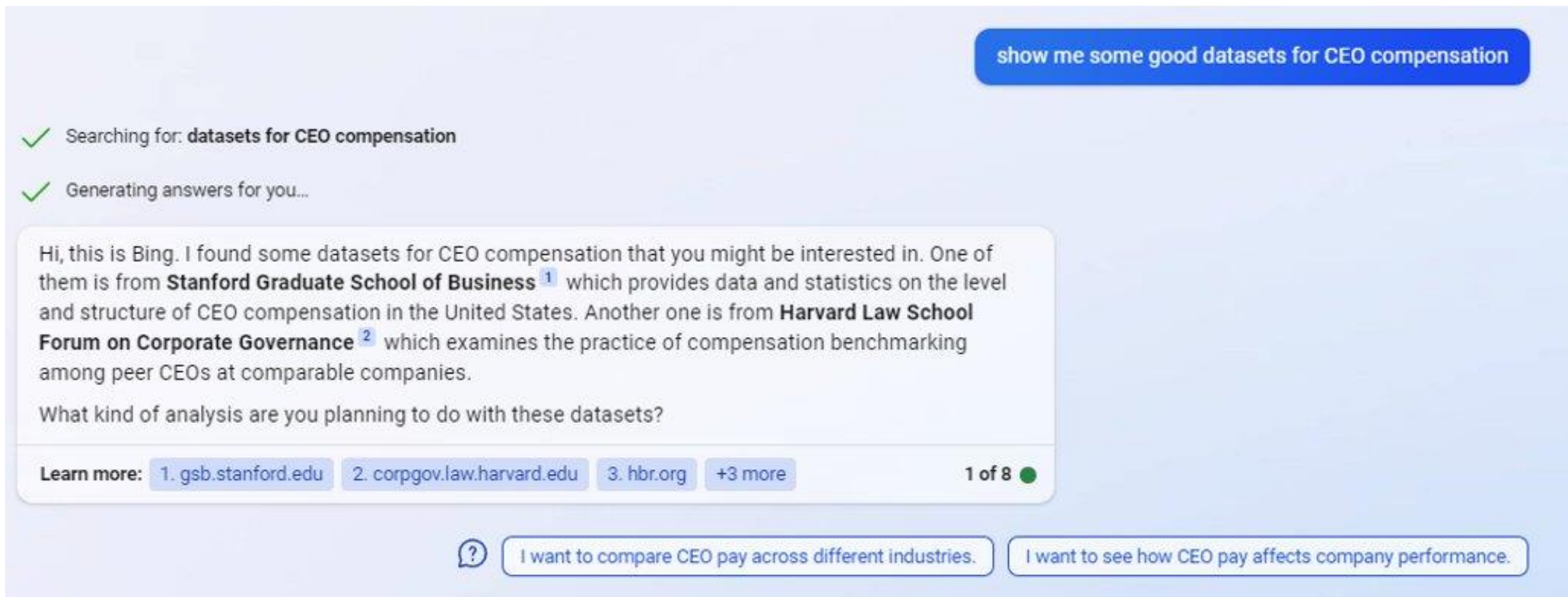
- Tuesdays & Thursdays: 8am to 9am
- Saturdays: 10am to 12pm

Students are to take their belonging with them. Unattended items will be removed.

Bing +GPT isn't very good in extracting data from tables

2. Implications for Reference and support services

Reasonably good chat bots will probably be used in libraries



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.

3. Implications for Open Access

- Will this strengthen the Open Access Citation advantage?

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

<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>

4. Cautions

- Technology is still very immature
 - Reliability of answers depends on training dataset (usually not known)
 - Use with caution, always verify. A citation does not mean it is correct
 - This is just the beginning, there is a lot of possible improvements coming

Be wary about putting in private, confidential information into ChatGPT or LLMs systems

Things I did not talk about

- Copyright issues
- Other Generative AI models like Text to Image generative AI models (Stable Diffusion, DALL-E2)
- Multimodal models (covers text, image, videos for input/output)
- Disinformation risks
- Detection of AI writing
- LLM's pattern matching or reasoning?

Impact on education & jobs

AI existential risk

Thank you – Any Questions?