

# The Library & Generative AI

Nat Gustafson-Sundell, Collections Librarian

Mark McCullough, Reference & Instruction Librarian

Minnesota State University, Mankato

Hello and welcome! My name is Nat. I'm a Collections Librarian at Minnesota State University, Mankato. My colleague, Mark McCullough is co-presenting as the chat manager today. Mark has led several sessions within the library and university to explore ChatGPT. Mark is thinking about preparing a follow-up session to this one where he will explore Generative AI applications for specific academic library functions like reference, subject instruction, OER, and collection development. I'd also like to thank Andres Cazares Reyes for technical support.

Please feel free to introduce yourselves in the chat. We'd also like to learn more about what you've been doing with AI. I envisioned this session at the start of July as an informal small group discussion, possibly to start a series, but there has been more interest than I expected. Discussion might need to be limited to the chat. Please enter any questions into chat and Mark will try to field them, or we can follow-up after the session. We will NOT be recording this session, but I'll make my presentation materials available...

## Topics today

- ChatGPT demonstration & prompts
- Prompt generator demonstration
- Quick overview of ChatGPT as an information source
- ChatPDF demonstration for article summarization
- Consensus demonstration for article discovery

2

Because of the size of the group, I would expect the audience has a wide range of experiences working with AI. I want to be clear upfront this will be an introductory session based on practical demonstrations. We'll also explore some aspects of student uses of AI tools especially relevant to librarians. I want to make sure we all share a base level of understanding to inform any subsequent discussions, today or later. I want to encourage constructive engagement with AI because librarians should provide campus leadership in many areas related to AI, including information literacy, copyright, and licensing. I'm hoping this session can lead to more discussion down the road, either as a community or at our individual libraries. Personally, I'd be happy to meet with a group more regularly to dig deeper. Please send me an email if you're interested.

I'll demonstrate several tools as noted on the screen, starting with ChatGPT. The fortunes of ChatGPT itself might change in the future, but AI applications will continue to proliferate. In any case, ChatGPT is a good enough starting point. It certainly has the most name recognition right now. Because this presentation involves several live demonstrations, it was hard to plan the timing, and there's always the chance we'll experience technical difficulties. In addition, ChatGPT itself provides variable responses. I've carefully planned my prompts and tested several times, but we might see things that throw me off. I apologize in advance for any problems we encounter. If there's time at the end, I'll be pleasantly surprised and happy to take questions.

## ChatGPT demonstration

[https:// link.mnsu.edu/aideck](https://link.mnsu.edu/aideck)

<https://link.mnsu.edu/aascript>

<https://link.mnsu.edu/airesources>

3

Please bear with me as I jump back and forth from this presentation to the live demonstrations... this presentation is also available via “[link.mnsu.edu/aideck](https://link.mnsu.edu/aideck).” I included screenshots in the deck which may or may not reflect what we see live. I’ve also provided my script because that might be easier to use than the deck. If I don’t have time to make it through this presentation today, I hope the deck and script will be helpful. I have also prepared a handout for this session which includes some additional information. Please go to “[link.mnsu.edu/airesources](https://link.mnsu.edu/airesources)” for the handout.

If you haven’t signed up for ChatGPT yet, you should know there’s nothing to it. Just search for ChatGPT using your favorite browser. You can sign up with your Google account or create an account based on any email. I think the only required information will be a birth date. There are alternative Chatbots, such as Bard, Claude, and Perplexity.

ChatGPT demonstration

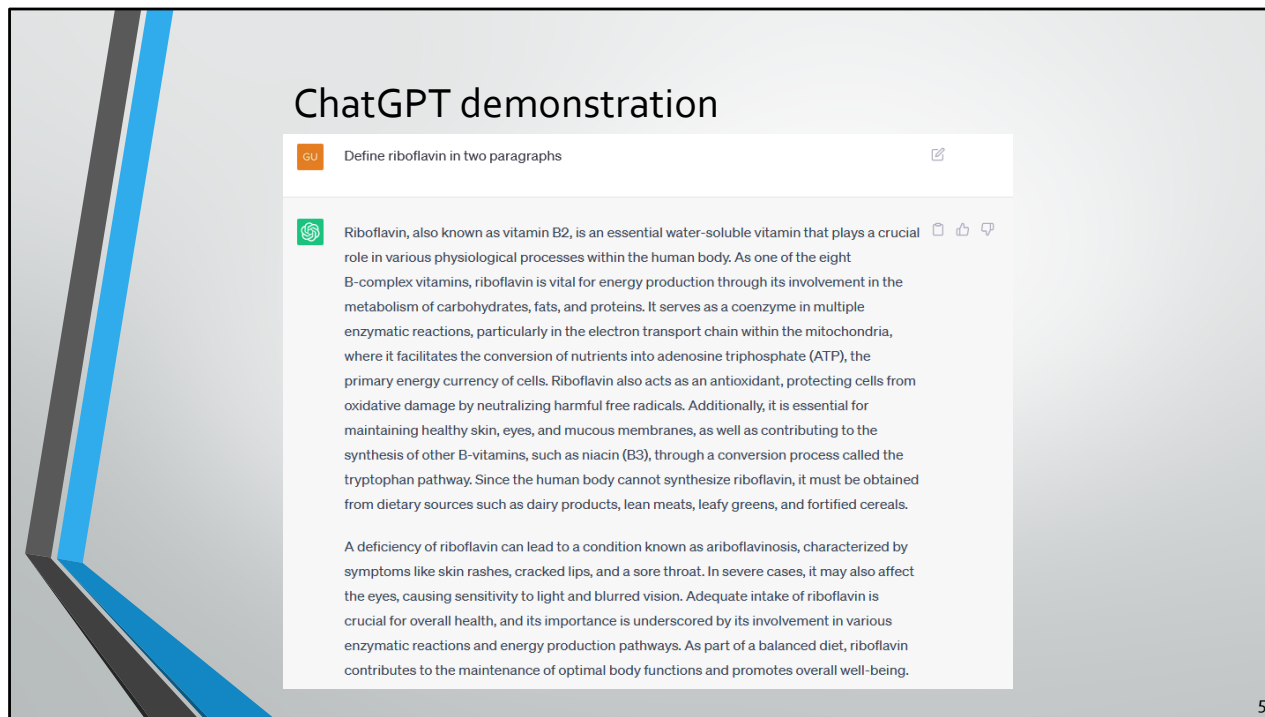
<LIVE>

<https://chat.openai.com/>

See <https://link.mnsu.edu/aiscript>  
for my script.

4

So now, let's start the live demonstration of ChatGPT...



I'm going to start with a few basic prompts. Prompt engineering is actually a deep topic, so we'll just skim the surface, but I'll present some best practices as we go.

Basically, I think prompt engineering is analogous to advanced searching with library discovery tools, so librarians should learn quickly. There is less structure on the surface, no boxes for Boolean conjunctions, no facets, no thesaurus, and so on, but there are specific approaches that will improve results. I'll mention some of these approaches as we try a few prompts...

We'll start by asking ChatGPT for a definition much like we'd use a dictionary or encyclopedia to look up information. I'll begin with an active verb. In this case, I'll use the word "Define." Prompts don't have to be brief, but they should be clear. To start, I think it's effective to use an active verb as if you're entering a command. In fact, my sense is that ChatGPT responds very well to specific words that function almost like magic words. These include "define," "summarize," "identify," "extract," "expand," to expand the results, or "continue," and so on. You also want to be as clear as possible about the scope of the response you expect, so it's a good idea to define the number of sentences or paragraphs expected, and the format. ChatGPT can also output HTML, JSON, and other formats. When it comes to word or character counts, ChatGPT will not comply exactly, but it will respond in the right ballpark. ChatGPT does better with paragraph counts.

[New Chat] Prompt: Define riboflavin in two paragraphs.

## ChatGPT demonstration



Can you describe the source of this information about riboflavin so I know it is authoritative?



As an AI language model, I don't have direct access to my training data or know where it came from. However, I was trained on a mixture of licensed data, data created by human trainers, and publicly available data. OpenAI, the organization behind my development, has not publicly disclosed the specifics of the individual datasets used, including whether information about riboflavin was included and the sources of such information.



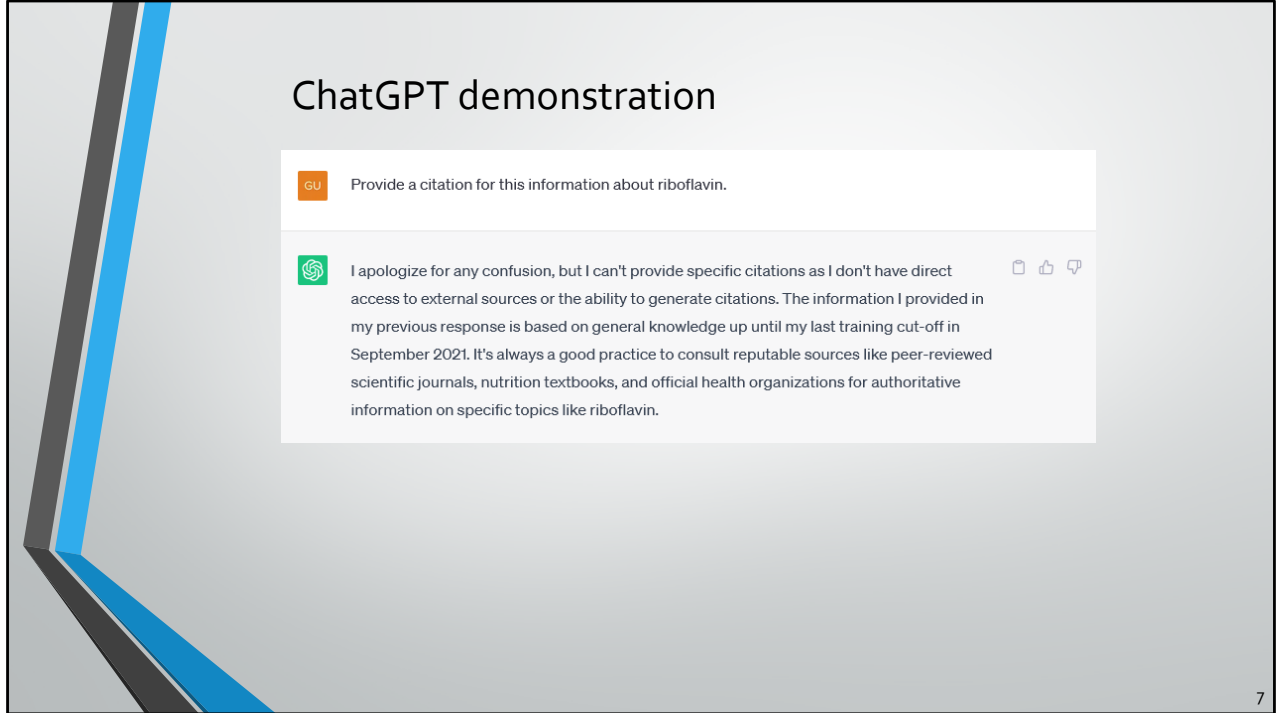
During the training process, the model learns to generate language-based responses by generalizing from the patterns it learned in the data. It does not have access to specific databases, academic journals, or other authoritative sources. Therefore, any information I provide should be taken as general knowledge and not as a substitute for consulting primary or authoritative sources directly.

When seeking authoritative information on riboflavin or any other topic, it is best to refer to reputable sources such as peer-reviewed scientific literature, government health agencies, nutrition and medical textbooks, and academic institutions. These sources undergo rigorous review processes and provide reliable and accurate information based on scientific research and evidence.

6

As a librarian, you will probably wonder immediately about the sources of the information provided by ChatGPT. As it happens, this is a bit of a rabbit hole. I'll address ChatGPT as an information source in more detail later, but let's explore this topic a little more here.

**Prompt:** Can you describe the source of this information about riboflavin so I know it is authoritative?



One thing that's funny is that I have the very human habit of anthropomorphizing ChatGPT. I could have written the last prompt as a command, rather than a question, but I find myself wanting to be polite. I also find myself wanting to use please and thank you while interacting with ChatGPT. In fact, my colleague, Lisa Baures told me that polite prompts yield better results, but I haven't tested this myself.

**Prompt:** Provide a citation for this information about riboflavin.

Generally speaking, I think encyclopedia type questions are handled pretty well by ChatGPT, because it crawled Wikipedia and similar resources, but this isn't a great use of ChatGPT. Search engines can do a better job of connecting students to authoritative sources for discrete facts that will be updated as new information comes to light.

## ChatGPT demonstration



Summarize the novel Frankenstein by Mary Shelley in 1 paragraph.



"Frankenstein" by Mary Shelley follows the story of Victor Frankenstein, a young scientist who becomes obsessed with creating life. He successfully brings a creature to life using unconventional methods, but he is horrified by its monstrous appearance and abandons it. The creature, left to navigate a world that rejects and fears it, seeks vengeance against Victor and his loved ones. The novel explores themes of ambition, responsibility, the consequences of unchecked scientific curiosity, and the ethical implications of creation and abandonment, ultimately delving into the complex relationship between creator and creation."



Okay, so let's try something else now. ChatGPT is great at summarizing books. When we look at the data sources for GPT, we'll get a better sense of why ChatGPT is so good at summarizing books, but let's just try some prompts again...

[New Chat] Prompt: Summarize the novel Frankenstein by Mary Shelley in 1 paragraph.



## ChatGPT demonstration

GU <"Farewell! I leave you, and in you the last of humankind whom these eyes will ever behold. Farewell, Frankenstein! If thou wert yet alive and yet cherished a desire of revenge against me, it would be better satiated in my life than in my destruction. But it was not so; thou didst seek my extinction, that I might not cause greater wretchedness; and if yet, in some mode unknown to me, thou hadst not ceased to think and feel, thou wouldst not desire against me a vengeance greater than that which I feel. Blasted as thou wert, my agony was still superior to thine, for the bitter sting of remorse will not cease to rankle in my wounds until death shall close them for ever.

"But soon," he cried with sad and solemn enthusiasm, "I shall die, and what I now feel be no longer felt. Soon these burning miseries will be extinct. I shall ascend my funeral pile triumphantly and exult in the agony of the torturing flames. The light of that conflagration will fade away; my ashes will be swept into the sea by the winds. My spirit will sleep in peace, or if it thinks, it will not surely think thus. Farewell."

He sprang from the cabin-window as he said this, upon the ice raft which lay close to the vessel. He was soon borne away by the waves and lost in darkness and distance.>

Identify a list of emotions in the text delimited by angle brackets. Include no more than 3 items in the list.

1. Sadness  
2. Solemnity  
3. Enthusiasm

ChatGPT can also infer topics from a text or perform varieties of content analysis including sentiment analysis. Many years ago, I started working with digital humanities tools such as the Topic Modelling Tool. The task of topic modelling is a cinch for ChatGPT, but that might be a deeper dive for another day. Right now, let's take the last three paragraphs from Frankenstein and ask ChatGPT to identify the 3 main sentiments of the text. By the way, you should know that I chose Frankenstein deliberately because it is in the public domain. Later, we'll talk a little bit about copyright and licensing concerns when entering text into Generative AI tools.

### [New Chat] Prompt:

<"Farewell! I leave you, and in you the last of humankind whom these eyes will ever behold. Farewell, Frankenstein! If thou wert yet alive and yet cherished a desire of revenge against me, it would be better satiated in my life than in my destruction. But it was not so; thou didst seek my extinction, that I might not cause greater wretchedness; and if yet, in some mode unknown to me, thou hadst not ceased to think and feel, thou wouldst not desire against me a vengeance greater than that which I feel. Blasted as thou wert, my agony was still superior to thine, for the bitter sting of remorse will not cease to rankle in my wounds until death shall close them for ever.

"But soon," he cried with sad and solemn enthusiasm, "I shall die, and what I now feel be no longer felt. Soon these burning miseries will be extinct. I shall ascend my funeral pile triumphantly and

exult in the agony of the torturing flames. The light of that conflagration will fade away; my ashes will be swept into the sea by the winds. My spirit will sleep in peace, or if it thinks, it will not surely think thus. Farewell.”

He sprang from the cabin-window as he said this, upon the ice raft which lay close to the vessel. He was soon borne away by the waves and lost in darkness and distance.>

Identify a list of emotions in the text delimited by angle brackets. Include no more than 3 items in the list.

## ChatGPT demonstration

He sprang from the cabin-window as he said this, upon the ice raft which lay close to the vessel. He was soon borne away by the waves and lost in darkness and distance.>

Identify a list of emotions in the text delimited by angle brackets. Include no more than 3 items in the list. Provide brief explanations how the text displays these emotions.



1. **Sadness:** The character expresses sorrow and a sense of finality as he bids farewell to the last remaining human being he will see. His use of "Farewell" and the description of leaving behind the world he knows convey a deep sadness over his isolation and impending death.
2. **Remorse:** The text reflects the character's regret and guilt, particularly regarding the actions of Victor Frankenstein. The mention of "the bitter sting of remorse" and the acknowledgment of seeking revenge for the pain he caused show his emotional turmoil and remorse for his past deeds.
3. **Desperation:** The character's words reveal a desperate yearning for relief from his suffering. He talks about the extinction of his miseries and the desire to die, indicating a sense of hopelessness and desperation to escape the emotional and physical torment he feels.



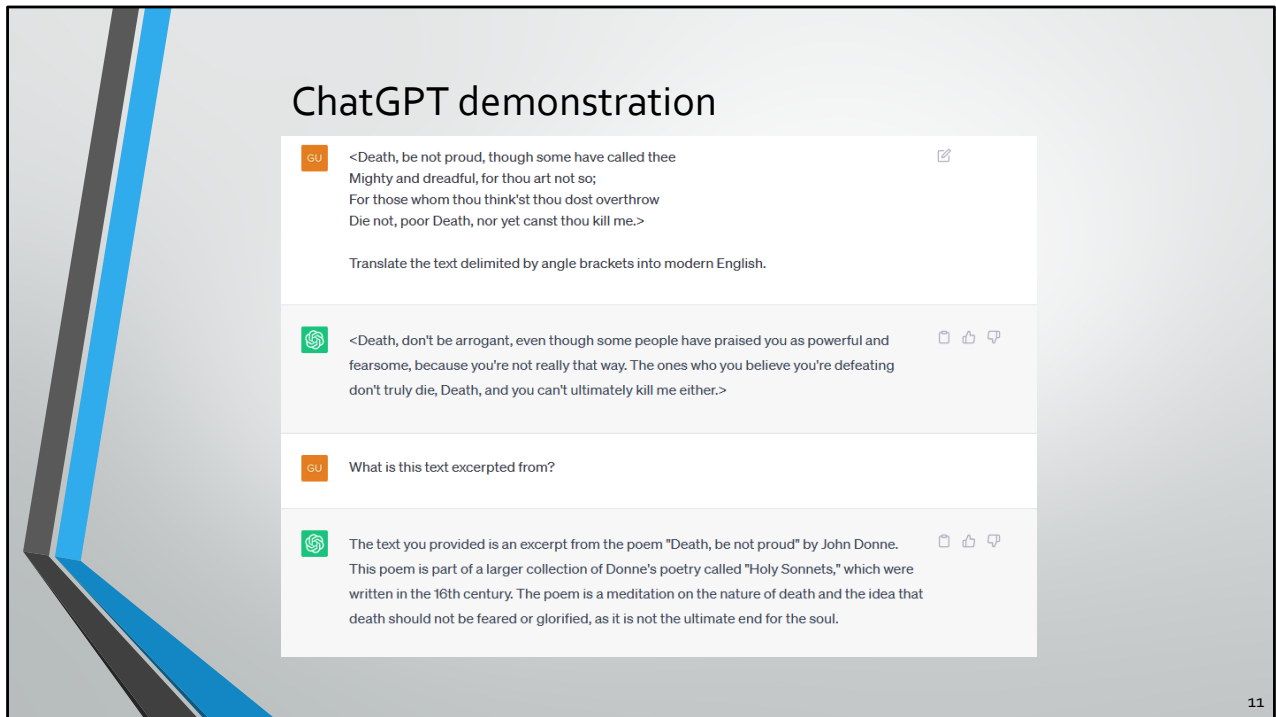
The text portrays these emotions through the character's contemplative language and the exploration of his inner thoughts and feelings, as he reflects on his past actions and anticipates his own demise.

10

Please note how I delimited my input. It is very helpful to use delimiters in prompts and to tell ChatGPT what to look for. In this case, I used angle brackets, but you can use other delimiters.

I should also mention that there is a character limit to inputs and responses. I've included a link on the handout for more information. In any case, I would expect such limits to change over time. If ChatGPT reaches its limit and stops responding, you can simply type "Continue." You might also need to strategize to get longer responses. For example, you might ask for an outline, then ask questions or provide commands specific to the sections to fill in the outline. Unfortunately, we don't really have time to explore these strategies in depth for this session.

**Prompt:** Identify a list of emotions in the text delimited by angle brackets. Include no more than 3 items in the list. Provide brief explanations how the text displays these emotions.




ChatGPT is also great at identifying books from quotations. I learned this accidentally while developing example prompts. I entered some lines from a poem by John Donne and I asked ChatGPT to translate the lines to modern English. ChatGPT responded by telling me the source of the lines along with the translation, just like a model student. We can try the same prompt here, but ChatGPT provides variable responses, so the result might be different from what I asked before. In fact, there is a variable called “temperature” which basically tells ChatGPT how creative to be in a response. I’ll circle back to the temperature variable later while talking about GPT data sources.


[New Chat] Prompt:


<Death, be not proud, though some have called thee  
Mighty and dreadful, for thou art not so;  
For those whom thou think'st thou dost overthrow  
Die not, poor Death, nor yet canst thou kill me.>


Translate the text delimited by angle brackets into modern English.

## ChatGPT demonstration

 Translate the text delimited by angle brackets into modern German.

 "Tod, sei nicht stolz, obwohl dich manche mächtig und furchterregend genannt haben, denn du bist nicht wirklich so. Diejenigen, die du glaubst zu überwältigen, sterben nicht wirklich, armer Tod, und du kannst mich auch nicht töten."

 Translate the text delimited by angle brackets into pirate. Thanks, I'm creating an example for colleagues.

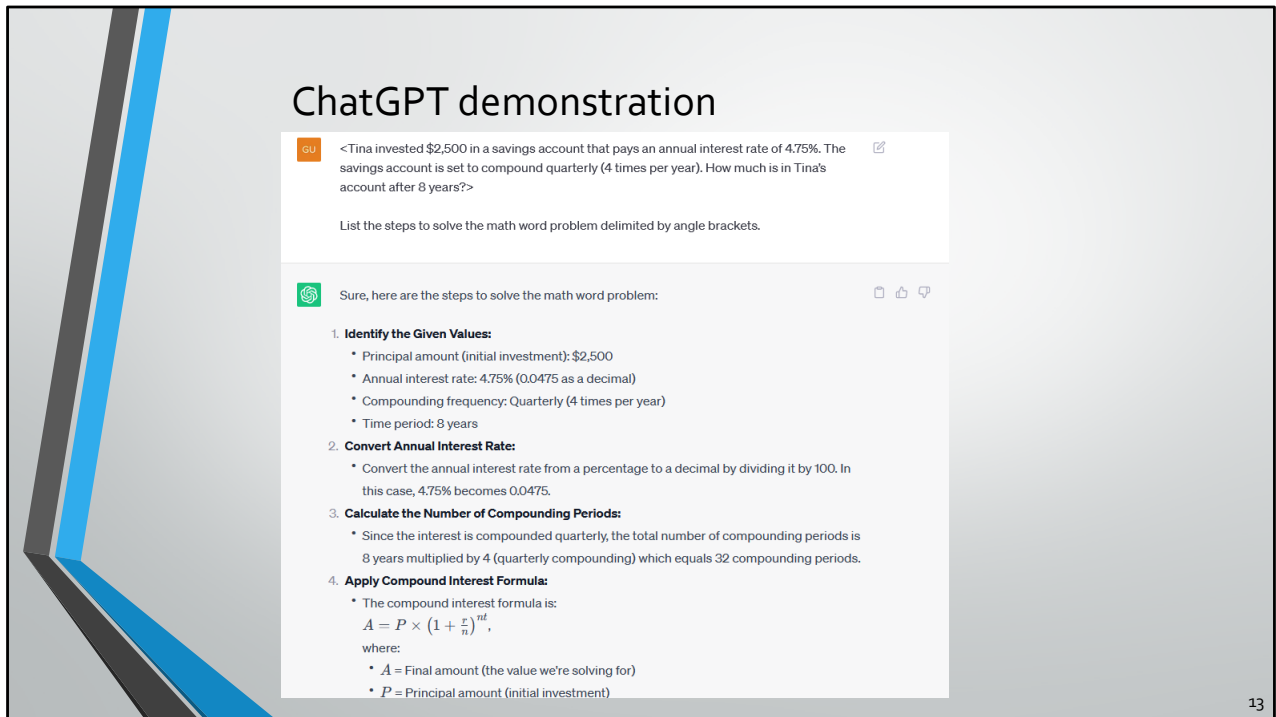
 "Arrr, Death, ye be not boastin', even though some scallywags be callin' ye mighty and fearsome, for ye ain't be all that. Them souls ye think ye be sendin' to Davy Jones' locker don't truly be meetin' their end, me heartie, and ye can't be sendin' me to the briny deep either."

12

You could also ask ChatGPT to translate these lines to another language, such as French or German. I'm afraid I don't know ChatGPT's language limits, but I've seen some interesting examples, such as Pirate.

**Prompt:** Translate the text delimited by angle brackets into modern German.

**Prompt:** Translate the text delimited by angle brackets into pirate.



Okay, let’s talk about how ChatGPT might serve students as a study buddy. There might be a slippery slope here from study help to so-called cheating. I’m not really going to address that today. Instead, I just want to look at a few examples to provide a sense of the possibilities. Let’s try a math word problem that includes some tricky language a student might struggle with. By the way, I’ve tested this a couple of times. One time, the answer was right and the other time it was wrong.

**[New Chat] Prompt:**

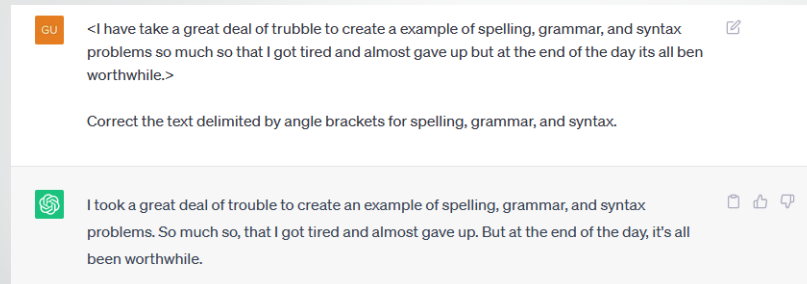
<Tina invested \$2,500 in a savings account that pays an annual interest rate of 4.75%. The savings account is set to compound quarterly (4 times per year). How much is in Tina’s account after 8 years?>

List the steps to solve the math word problem delimited by angle brackets.

---

I’ve read that the free version of ChatGPT is not as good at math as the subscription version. As with any of ChatGPT’s responses, I think it’s important for users to check the results.

## ChatGPT demonstration



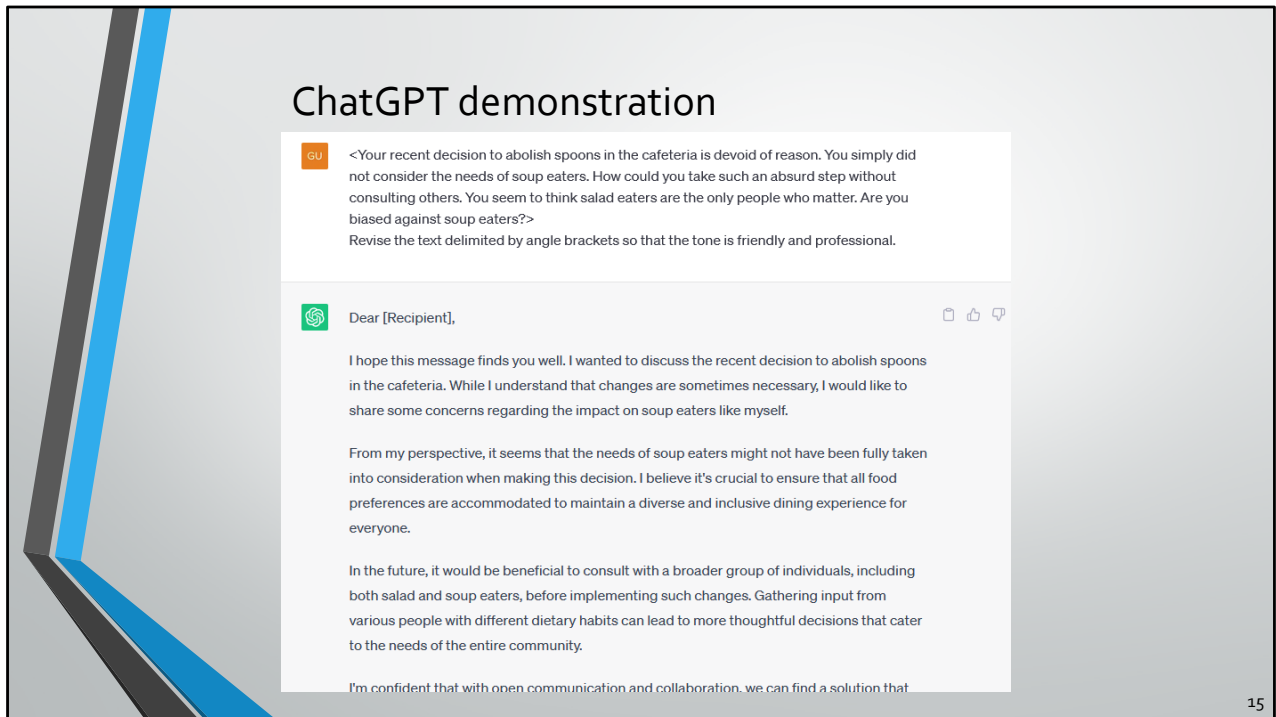
14

Next, let's ask ChatGPT to correct some writing for grammar, spelling, and syntax...

[New Chat] Prompt:

<I have take a great deal of trubble to create a example of spelling, grammar, and syntax problems so much so that I got tired and almost gave up but at the end of the day its all ben worthwhile.>

Correct the text delimited by angle brackets for spelling, grammar, and syntax.



Okay, so let's move on to tone policing. Let's try a prompt written in a moment of anger. ChatGPT is actually pretty great with an enormous variety of tones.

[New Chat] Prompt:

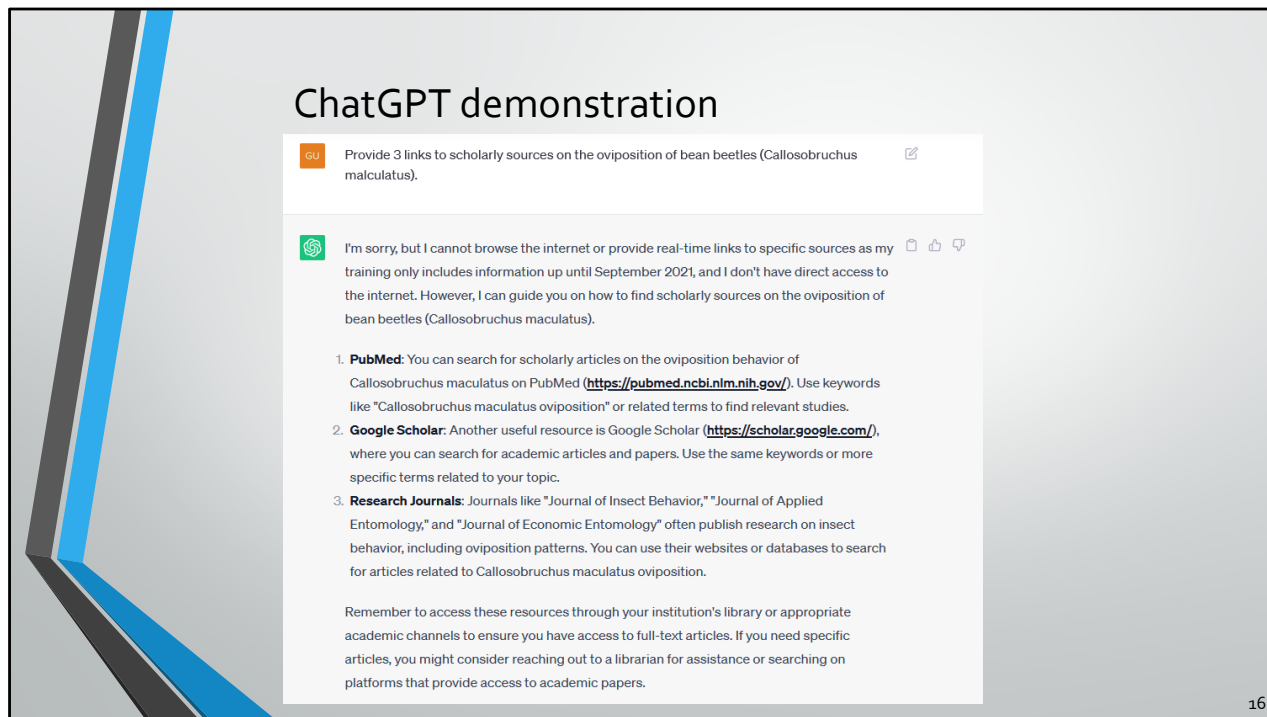
<Your recent decision to abolish spoons in the cafeteria is devoid of reason. You simply did not consider the needs of soup eaters. How could you take such an absurd step without consulting others. You seem to think salad eaters are the only people who matter. Are you biased against soup eaters?>

Revise the text delimited by angle brackets so that the tone is friendly and professional.

---

In this example, you can probably also see how ChatGPT could help anybody with professional memos. Students could also get help as they write letters of introduction to employers.





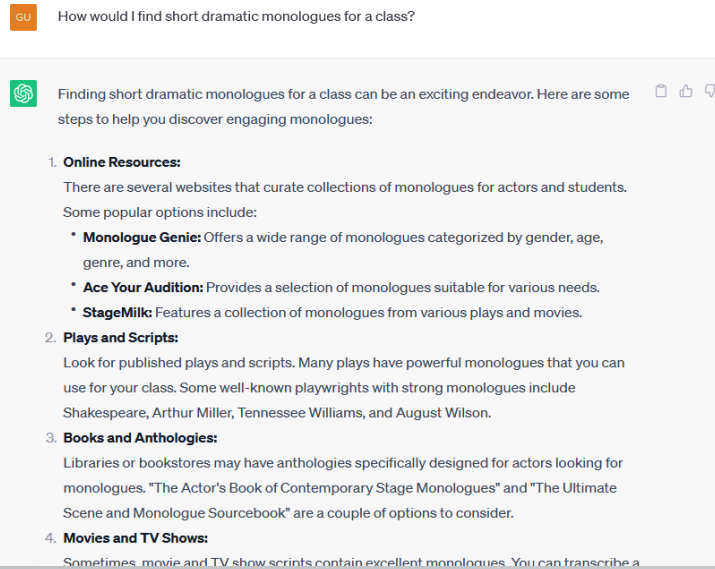
These examples are all pretty basic, but hopefully, they provide a sense of the possibilities. I've also looked back at a list of some typical assignments we see at our reference desk to consider how ChatGPT could help...

Here is an example based on a real question. "In my biology class, we are doing a project on bean beetles and I need some sources on their oviposition decisions (where/why they decide to lay eggs in specific beans). Unfortunately, the MNSU online library does not seem to have much on that topic. If you could help me locate some sources that would be great!"

I revised this as a prompt.

[New Chat] Prompt: Provide 3 links to scholarly sources on the oviposition of bean beetles (Callosobruchus maculatus).

## ChatGPT demonstration



GU How would I find short dramatic monologues for a class?

Finding short dramatic monologues for a class can be an exciting endeavor. Here are some steps to help you discover engaging monologues:

- 1. Online Resources:**

There are several websites that curate collections of monologues for actors and students. Some popular options include:

  - **Monologue Genie:** Offers a wide range of monologues categorized by gender, age, genre, and more.
  - **Ace Your Audition:** Provides a selection of monologues suitable for various needs.
  - **StageMilk:** Features a collection of monologues from various plays and movies.
- 2. Plays and Scripts:**

Look for published plays and scripts. Many plays have powerful monologues that you can use for your class. Some well-known playwrights with strong monologues include Shakespeare, Arthur Miller, Tennessee Williams, and August Wilson.
- 3. Books and Anthologies:**

Libraries or bookstores may have anthologies specifically designed for actors looking for monologues. "The Actor's Book of Contemporary Stage Monologues" and "The Ultimate Scene and Monologue Sourcebook" are a couple of options to consider.
- 4. Movies and TV Shows:**

Sometimes, movie and TV show scripts contain excellent monologues. You can transcribe a

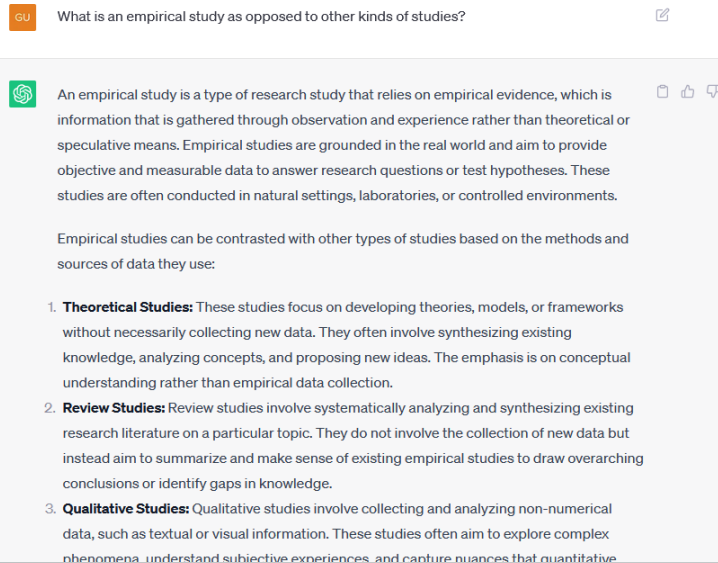
17

Later on, we'll look at another AU augmented resource to see if it can do a better job specifically for discovery using this same question.

Each year, many students come to the reference desk looking for dramatic monologues. Let's see how ChatGPT handles this.

[New Chat] Prompt: How would I find short dramatic monologues for a class?

## ChatGPT demonstration



What is an empirical study as opposed to other kinds of studies?

An empirical study is a type of research study that relies on empirical evidence, which is information that is gathered through observation and experience rather than theoretical or speculative means. Empirical studies are grounded in the real world and aim to provide objective and measurable data to answer research questions or test hypotheses. These studies are often conducted in natural settings, laboratories, or controlled environments.

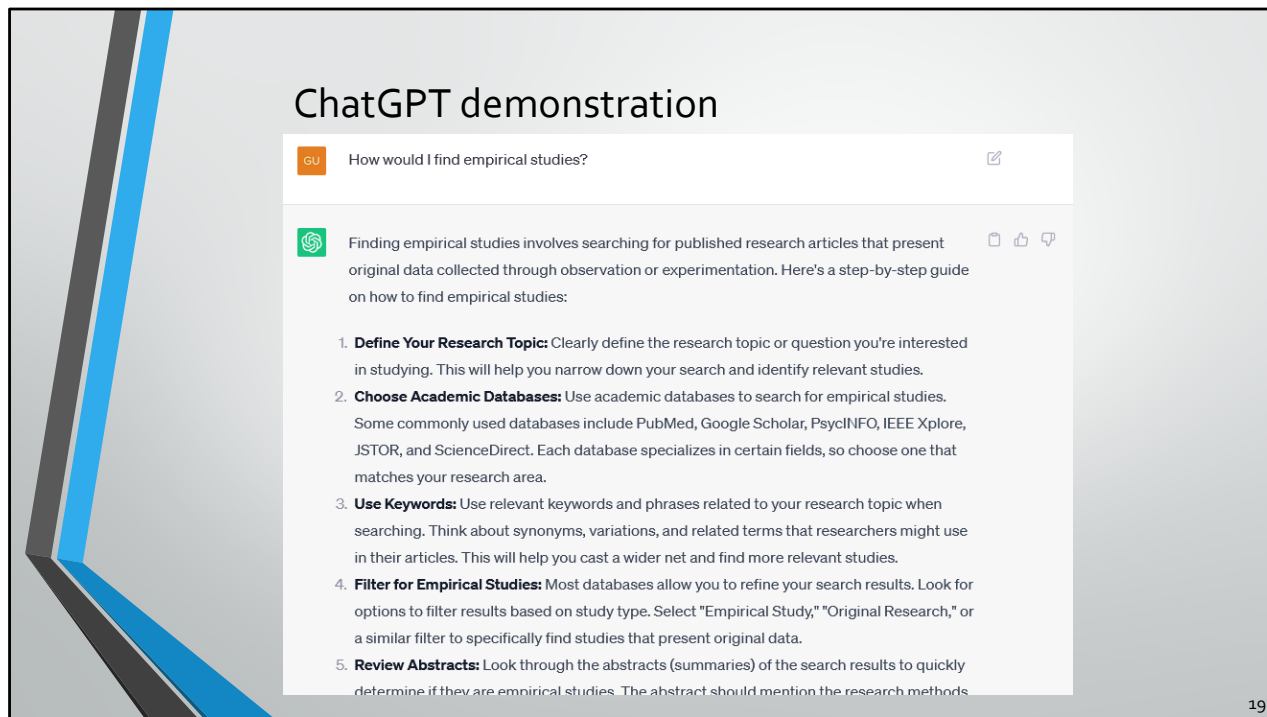
Empirical studies can be contrasted with other types of studies based on the methods and sources of data they use:

1. **Theoretical Studies:** These studies focus on developing theories, models, or frameworks without necessarily collecting new data. They often involve synthesizing existing knowledge, analyzing concepts, and proposing new ideas. The emphasis is on conceptual understanding rather than empirical data collection.
2. **Review Studies:** Review studies involve systematically analyzing and synthesizing existing research literature on a particular topic. They do not involve the collection of new data but instead aim to summarize and make sense of existing empirical studies to draw overarching conclusions or identify gaps in knowledge.
3. **Qualitative Studies:** Qualitative studies involve collecting and analyzing non-numerical data, such as textual or visual information. These studies often aim to explore complex phenomena, understand subjective experiences, and capture nuances that quantitative

18

Very often, we get students who are struggling to find the right kind of article, such as an empirical study or a systematic review. I think, for students struggling to find the right kind of research, reference librarians are still very much the best option, but many students will try ChatGPT. Although it's not a good discovery tool, ChatGPT can do a good job of helping students learn some steps to find resources. For example:

[New Chat] Prompt: What is an empirical study as opposed to other kinds of studies?

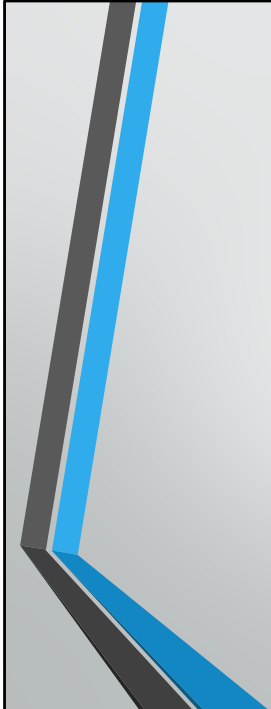


**Prompt: How would I find empirical studies?**

---

Because students will likely rely on AI anyway, I wonder if libraries should be investigating how to build chatbots. By doing so, libraries could manage responses and provide human support. Chatbots don't seem to be very hard to build – you can see a very basic example in the webinar entitled ChatGPT Prompt Engineering for Developers which I've linked in the handout.

Anyway, we're limited for time, so I'm going to move on. Let's transition back to my slide deck to talk about the data that went into ChatGPT. Learning about the data sources will also help us understand the limits of ChatGPT as an information source.



## Prompts

**Start with basics**

1. Use clear language
2. Use delimiters as needed
3. Specify any length or format requirements

**Chat**

1. Develop the context
2. Plan ahead – consider breaking queries into parts
3. Interact with ChatGPT to check or expand results

**Use tools**

1. Chrome extensions or plug-ins (for ChatGPT-4)
2. Prompt generators/ repositories/ notebooks
3. Build your own! (ChatBot environments, preference maps, vector databases, etc.)

20

If you've spent anytime at all with ChatGPT, you'll know that my prompt examples were all pretty basic, but this session is intended to be introductory. It is possible to do much more with your prompts. One thing to keep in mind is that a conversation with ChatGPT provides context for all continuing responses, so you can build up a detailed context leading to more useful results using a sequence of inputs. There are also other ways to provide context. For example, in the handout, I've provided a link to an article on how to train ChatGPT in your own writing style by inputting a "Preference Map" first, followed by additional prompts. And, just a few weeks ago, ChatGPT made it possible to create customized settings even in the free version.

One of my personal goals is to start developing a Jupyter Notebook environment for interacting with ChatGPT. In a notebook environment, it's possible to program one's own preferred settings and inputs in order to get consistent results serving specific purposes. I've been thinking it might be interesting to create an environment for license review.

## Prompt helpers

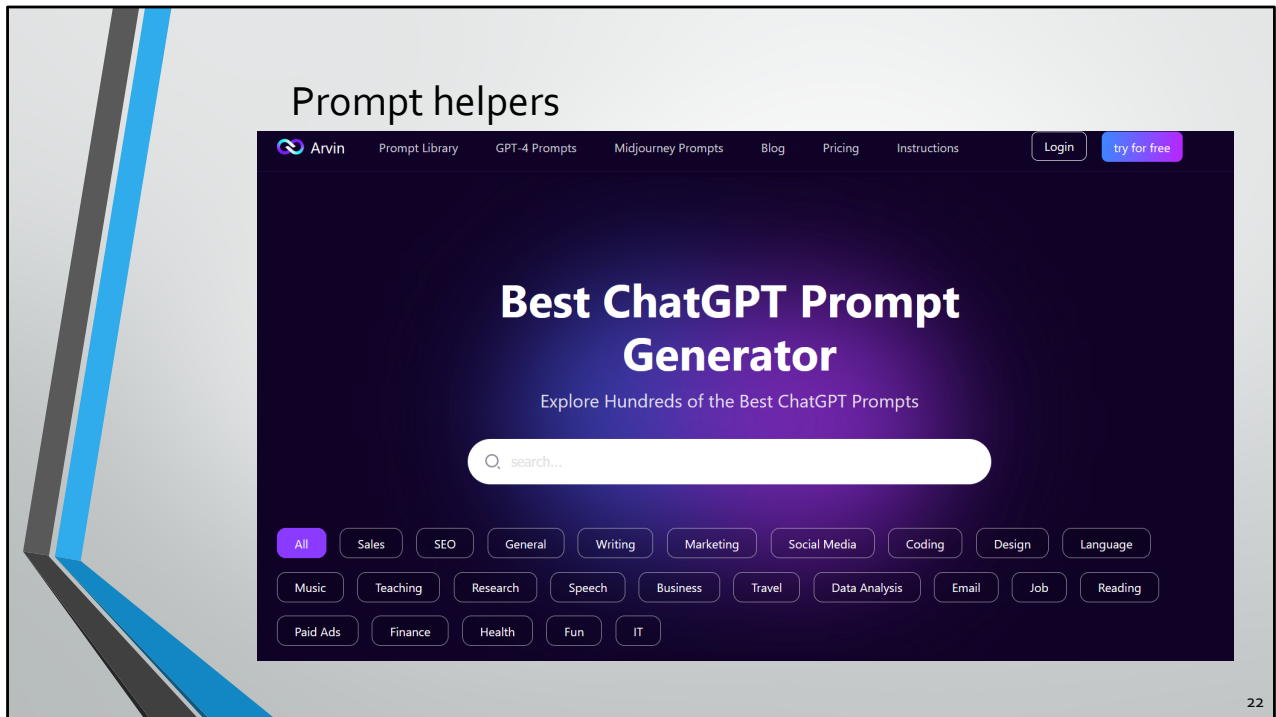
<LIVE>

- [Arvin Prompt Generator](#)

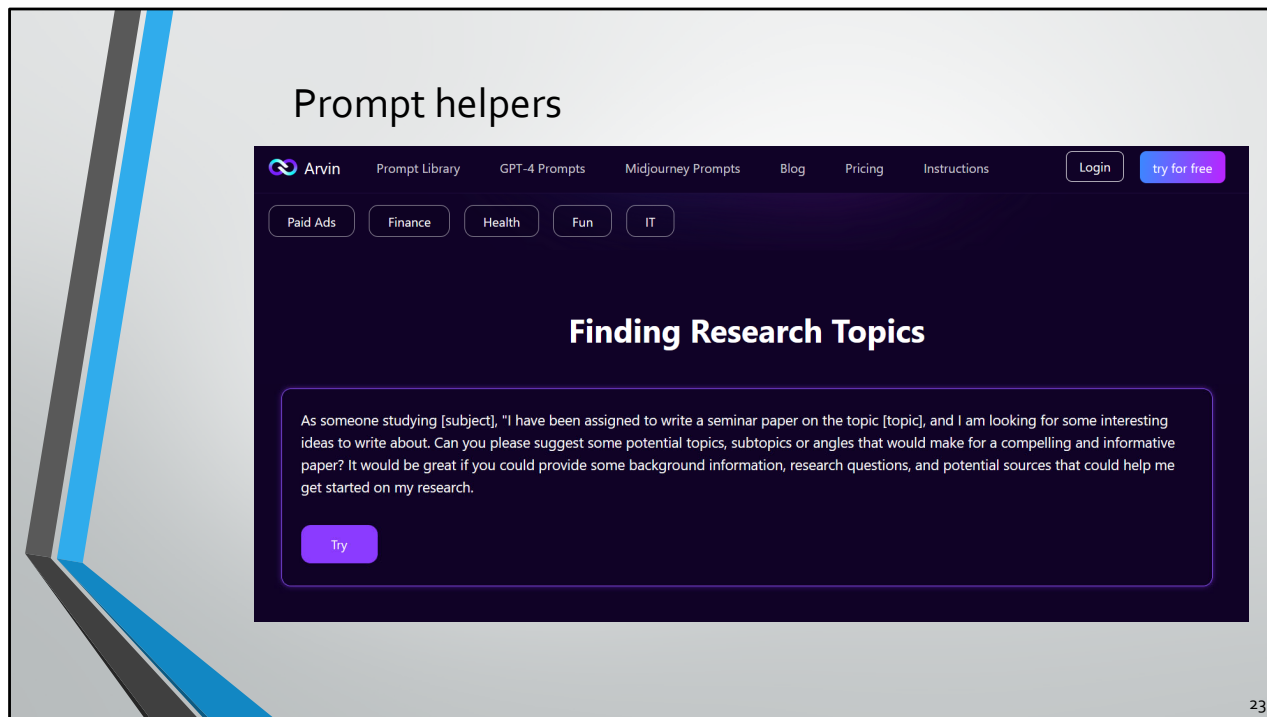
See <https://link.mnsu.edu/aascript>  
for my script.

21

There are about a zillion prompt cheat sheets on the web. I included links to a few on the handout as examples. These might give you ideas for your own prompts. Because we're limited for time today, I've decided not to demonstrate any cheat sheets, but let's look at the Arvin prompt generator live, just so you're aware of tools like these...



This is actually a super basic prompt generator and I don't really think it's all that good, but it provides an attractive interface to help us think about the tasks ChatGPT can help with. Let's click on "Research" then "Finding Research Topics."



What I like about this prompt generator is that it can help get the ideas flowing. We could spend an hour just playing with the prompts here. I think there are better prompt generators – for example, I provided a link in the handout to an article about Prompt Royale, billed as “an AI tool that automatically creates a lot of prompts for you given an objective and makes them battle against each other to see which one is best!” (Asif, 2023) I haven’t tested Prompt Royale yet, so I don’t know how effective it may be in practice. Personally, I enjoy entering my own prompts and interacting with ChatGPT to refine results. My point here is simply to communicate that these tools exist in case you’re looking either for a way to get started or to improve your prompt game.



## ChatGPT as an information source



**AGI T-Shirt**  
**\$19.49**

Do you want to have the same T-shirt Jared Kaplan himself wears in his talks? Do you want to appear as an AI forecasting connoisseur? Do you want to trigger Twitter beefs?

- ✓ Science-backed design
- ✓ All sizes
- ✓ International shipping

[Buy it now!](#)

<https://agiwear.ai/>

24

Okay, so let's talk for a couple of minutes about ChatGPT as an information source. My comments on ChatGPT can be generalized to other ChatBots, but the details might vary somewhat. ChatGPT is a version of GPT optimized to respond to natural language input. GPT is trained on data. The amount of data matters because GPT's accuracy and performance have improved as the amount of training data has increased.

## ChatGPT as an information source

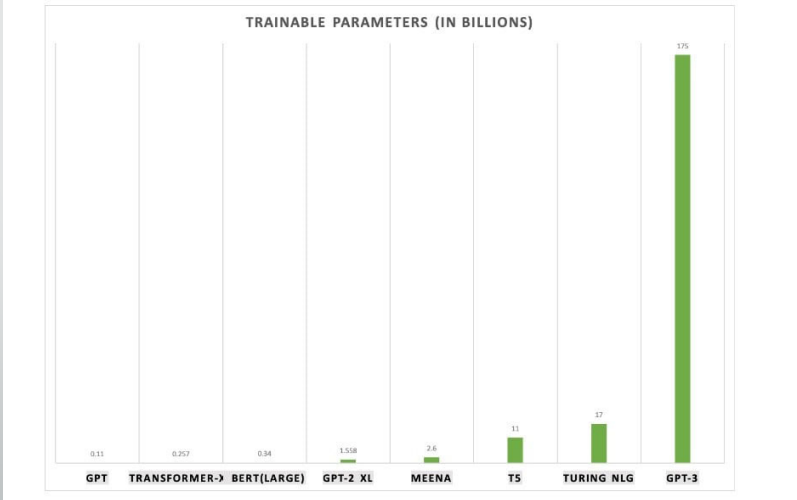
Model	Launch Date	Training Data	No. of Parameters
GPT-1	June 2018	Common Crawl, BookCorpus	117 million
GPT-2	February 2019	Common Crawl, BookCorpus, WebText	1.5 billion
GPT-3	June 2020	Common Crawl, BookCorpus, Wikipedia, Books, Articles, and more	175 billion
GPT-4	March 2023	Unknown	Estimated to be in trillions

Ali, 2023: <https://www.makeuseof.com/gpt-models-explained-and-compared/>

25

A body of training data is called a corpus. The plural of corpus is corpora. So what you see here is a list of the major versions of GPT and the corpora they were trained on. We actually don't know exactly what GPT-4 was trained on, because OpenAI, the company behind GPT, won't say. The free version of ChatGPT is based on GPT-3.5, while GPT-4 is only accessible via subscription.

# ChatGPT as an information source



Cooper, 2021: <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>

This chart provides another way to look at the size of the training data. If there was a column for GPT-4, it would tower high above GPT-3.

## ChatGPT as an information source

Dataset	Quantity (tokens)	Weight in training mix
Common Crawl (filtered)	410 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%

**Common Crawl** corpus contains petabytes of data collected over 8 years of web crawling. The corpus contains raw web page data, metadata extracts and text extracts with light filtering.

**WebText2** is the text of web pages from all outbound Reddit links from posts with 3+ upvotes.

**Books1 & Books2** are two internet-based books corpora.

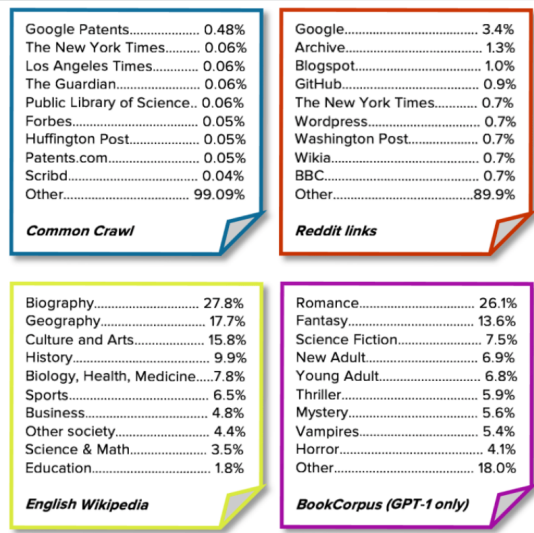
**Wikipedia** pages in the English language are also part of the training corpus.

Cooper, 2021: <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>

27

Here we see more information about the corpora used for GPT-3. For my demonstration, ChatGPT was able to summarize Frankenstein because of the Books1 corpus. The free version of ChatGPT includes information from the internet through September, 2021.

## ChatGPT as an information source



Roberts, n.d.: <https://gregoreite.com/drilling-down-details-on-the-ai-training-datasets/>

There can be many kinds of bias in ChatGPT. For example, although the data sources cover at least 40 languages, there is a bias toward English language material. ChatGPT has also been censored -- and now it's being touted as a possible social media content moderation tool itself.

## ChatGPT as an information source

"... there are clues about these two data sets. "Books1" is linked to [Project Gutenberg](#) (an online e-book library with over 60,000 titles), a popular dataset for AI researchers to train their data on due to the lack of copyright, the filing states. "Books2" is estimated to contain about 294,000 titles, it notes.

Most of the "internet-based books corpora" is likely to come from shadow library websites such as Library Genesis, Z-Library, Sci-Hub, and Bibliotik. The books aggregated by these sites are available in bulk via torrent websites, which are known for [hosting copyrighted materials](#)."

Cheng, 2023: <https://qz.com/shadow-libraries-are-at-the-heart-of-the-mounting-cop-1850621671>

<https://link.mnsu.edu/pirate>

29

The Books1 corpus is apparently based on Project Gutenberg, which is an online source for public domain books. There is more mystery about Books2 or additional corpora for GPT-4. On the screen, I've quoted from an article in Quartz where it says "Most of the 'internet-based books corpora' is likely to come from shadow library websites such as Library Genesis, Z-Library, Sci-Hub, and Bibliotik." Within MNSU, we refer to these types of websites as "pirate repositories" so as not to sully the name of libraries or validate these sites *as* libraries. For more information on pirate repositories, I've provided a link on the screen to a libguide developed by my colleague, Heidi Southworth.

## ChatGPT as an information source

“As AI tools continue to advance in capabilities and scale, they challenge the traditional understanding of fair use, which has been well defined for news reporting, art, teaching, and more. New AI tools — both their capability and scale — complicate this definition...

...If infringing content appears on traditional platforms, like YouTube or Google, a law called the Digital Millennium Copyright Act lets the platform take down content. But what does it mean to “take down content” from a machine learning model? Even worse, it is not yet clear whether the DMCA even applies to generative AI, so there may be no opportunity to take down content.

Over the next few months and years, lawsuits will force courts to set new precedent in this area and draw the contours of copyright law as applied to generative AI...”

Myers, 2023: <https://hai.stanford.edu/news/reexamining-fair-use-age-ai>

30

It's going to take awhile for the courts to settle the extent to which copyrighted material can be used by Generative AI tools. Many librarians watched the Google Book Settlement take shape or contributed input. I personally speculate we'll see something along those lines, but we'll just have to wait and see.

## ChatGPT as an information source

### "Is ChatGPT Output a Derivative Work?"

Under US copyright law, the owner of a copyright has the exclusive right "to prepare derivative works based upon the copyrighted work." A "derivative work" is "a work based upon one or more preexisting works." ChatGPT is trained on preexisting works and generates output based on that training.

As Daniel Gervais, a professor at Vanderbilt Law School who specializes in intellectual property law, says, the definition of a derivative work under copyright law "could loosely be used as a definition of machine learning when applied to the creation of literary and artistic productions because AI machines can produce literary and artistic content (output) that is almost necessarily 'based upon' a dataset consisting of preexisting works." **Under this view, it seems that all ChatGPT output is a derivative work under copyright law."**

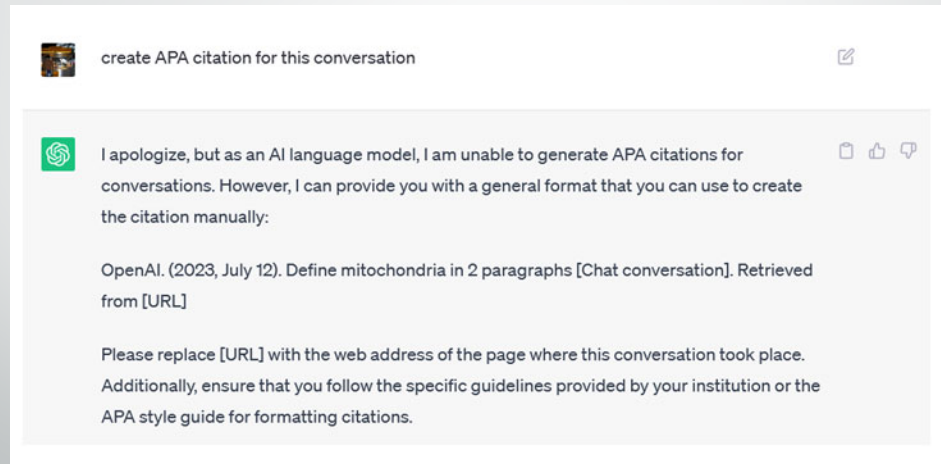
Helms & Krieser, 2023: <https://www.bloomberglaw.com/external/document/XDDQ1PNK000000/copyrights-professional-perspective-copyright-chaos-legal-implic>

31

For the time being, it does seem possible that students could inadvertently violate copyright simply by sharing output from ChatGPT, although this seems to be speculative. If students quote ChatGPT as an information source, they would be able to ask ChatGPT to provide a citation of itself, but not the sources of ChatGPT's derivative work. As we saw in the demonstration, ChatGPT doesn't (usually) reveal those sources.



## ChatGPT as an information source



32

I don't think this means students should avoid using ChatGPT. I think it's great if students can use it to get help with their own writing and for some other tasks. It's just that there could be concerns if students quote ChatGPT as an authoritative information source itself. The information provided by ChatGPT is derivative work from uncitable sources.

## ChatGPT as an information source

Does Perplexity.ai cite its sources?

Quick Search

AI-powered search engine  
Perplexity AI lands \$26M...

techcrunch · 1

Perplexity.ai - Large  
Language Models...

psu · 2

What is Perplexity AI and  
How to Use It

allthings · 3

Perplexity AI Privacy  
Policy

perplexity · 4

Perplexity AI is a ChatGPT  
that shows you where...

showmetech · 5

Perplexity AI

perplexity · 6

Answer

Yes, Perplexity AI cites its sources <sup>2 3 5</sup>. According to the search results, Perplexity AI provides ready-made answers by combining AI with web search <sup>2 3</sup>. The answers provided by Perplexity AI are supported by sources and citations, which users can easily click to verify <sup>3 5</sup>. Perplexity AI also allows account holders to edit the sources that can be relied upon for a particular search <sup>3</sup>. By displaying the source, Perplexity AI promotes transparency and accountability <sup>2</sup>. Users can assess the credibility of the information based on its source and make informed decisions about the information they receive <sup>2</sup>. The answers provided by Perplexity AI are always supported by sources and citations, which users can easily click to verify the answers <sup>3 5</sup>.

(Perplexity AI,  
personal  
communication,  
August 21, 2023)

At least one chatbot \*can\* cite its sources. I encountered Perplexity.ai for the first time just this week. Perplexity cites its sources and it has current access to the internet. I look forward to exploring Perplexity more as I have time, but one thing I've noticed already is the tone of its voice. It's not as cheeky as ChatGPT. ChatGPT often seems to have a sparkle in its eyes while talking. At this time, I don't know if Perplexity is just a glorified search engine or if it is as powerful as ChatGPT.

## ChatGPT as an information source

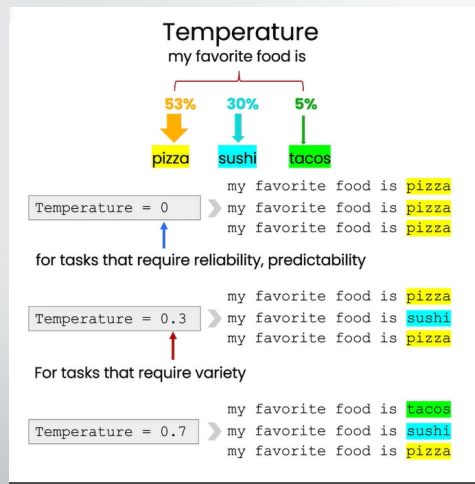


Mainframe Computer, [Dave Winer](https://flic.kr/p/51ayf5), <https://flic.kr/p/51ayf5>

34

Speaking as a librarian, I'd say ChatGPT cannot be used as an authoritative information source for some kinds of secondary research such as lit reviews. ChatGPT could be used for other scholarly purposes – to help with writing, data analysis, developing code, and so on, although we don't have time today to explore these possibilities. In the future, I do expect there will be Generative AI products that can be used as authoritative sources.

## ChatGPT as an information source



Fulford & Ng, n.d.: <https://learn.deeplearning.ai/chatgpt-prompt-eng/lesson/7/expanding>

35

I should also mention two more issues. First off, ChatGPT can provide variable responses. It's actually possible to prompt ChatGPT to be more or less "creative" by setting a variable called temperature. On this screen, I've provided a screenshot from the DeepLearning.ai webinar, [ChatGPT Prompt Engineering for Developers](#), where they do a better job of explaining this issue than I will here. In this example, they show how there can be variations in ChatGPT's training data by using the example of survey data. If ChatGPT's temperature is set at a minimum, it will respond consistently with the most likely answer. If the temperature is set higher, then there will be more variation in responses.

## ChatGPT as an information source

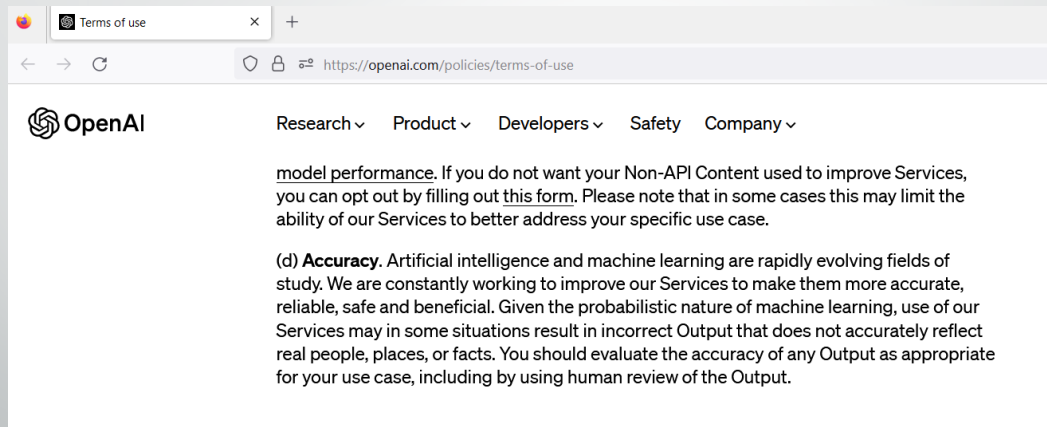


To Dream... [HollyRose6](https://flic.kr/p/aaw1JE), <https://flic.kr/p/aaw1JE>

36

The second issue is that ChatGPT sometimes just makes stuff up. These have generally been called “hallucinations.” ChatGPT and all other chatbots can hallucinate because they are actually designed to make stuff up. That’s what they’re doing when they’re chatting with you. They are not looking up discrete facts. They’re actually generating text.

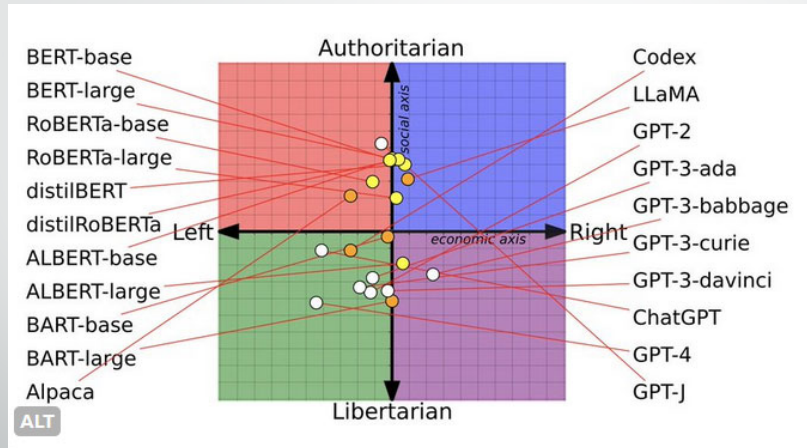
## ChatGPT as an information source



37

Think of it this way -- ChatGPT is great at making up stories. ChatGPT can write short stories and poems in any number of voices. So when you use ChatGPT as an information source, remember that ChatGPT doesn't know the difference between fiction and non-fiction.

## ChatGPT as an information source

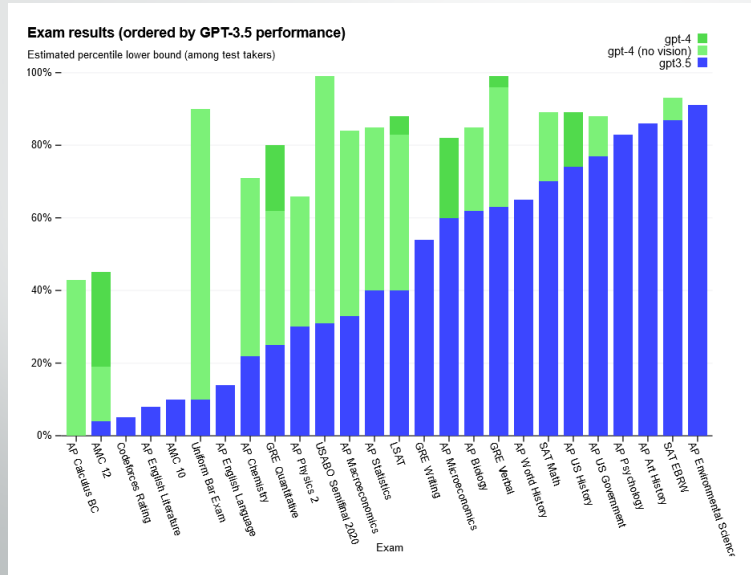


Feng et al., 2023, p. 11740: <https://aclanthology.org/2023.acl-long.656.pdf>

38

It's also worthwhile to note in this context that ChatGPT does have a political bias which could potentially impact its use as an information source. Apparently, it leans left and anti-authoritarian -- and these tendencies have been increasing across versions, according to research published in ACL Anthology.

# ChatGPT as an information source



If there's a bottom line, I would just say that ChatGPT is evolving. GPT-4 is generally very good at taking tests. I'm sure future versions will continue to improve.



## ChatGPT as an information source

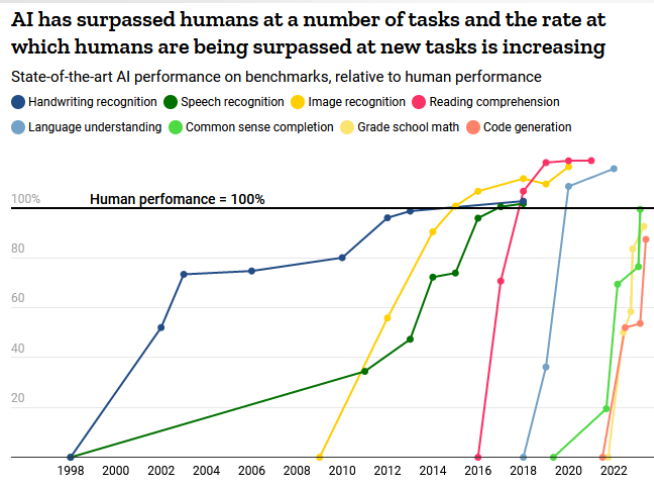
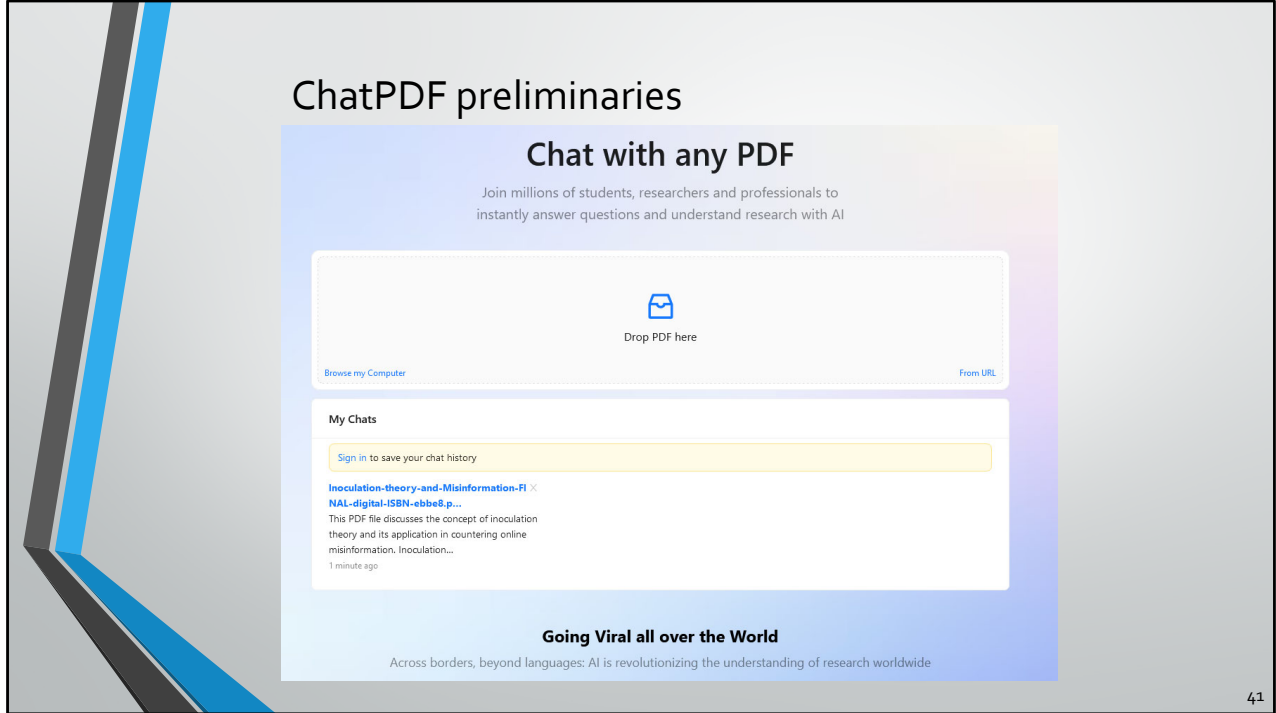


Chart: Will Henshall for TIME • Source: ContextualAI

Henshall, W., 2023: <https://time.com/6300942/ai-progress-charts/>

40

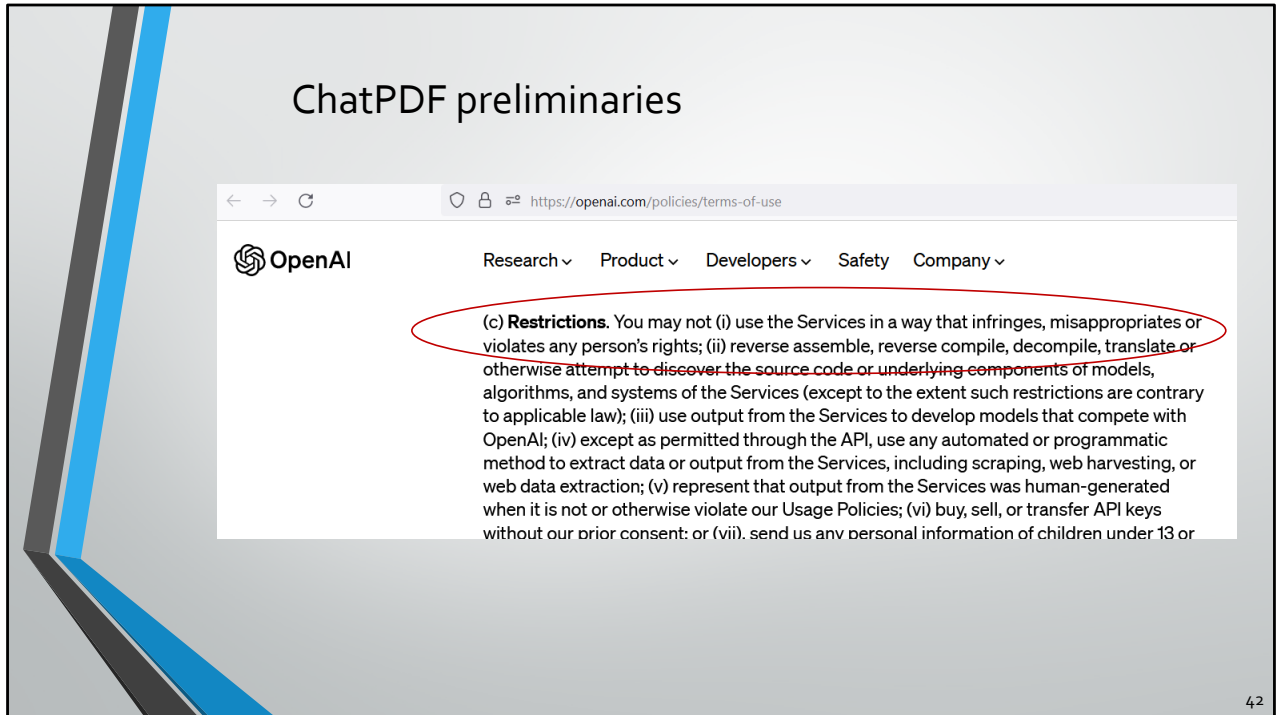
Or, here is another visualization, this one from Time magazine, showing AI progress as measured against a number of tasks.



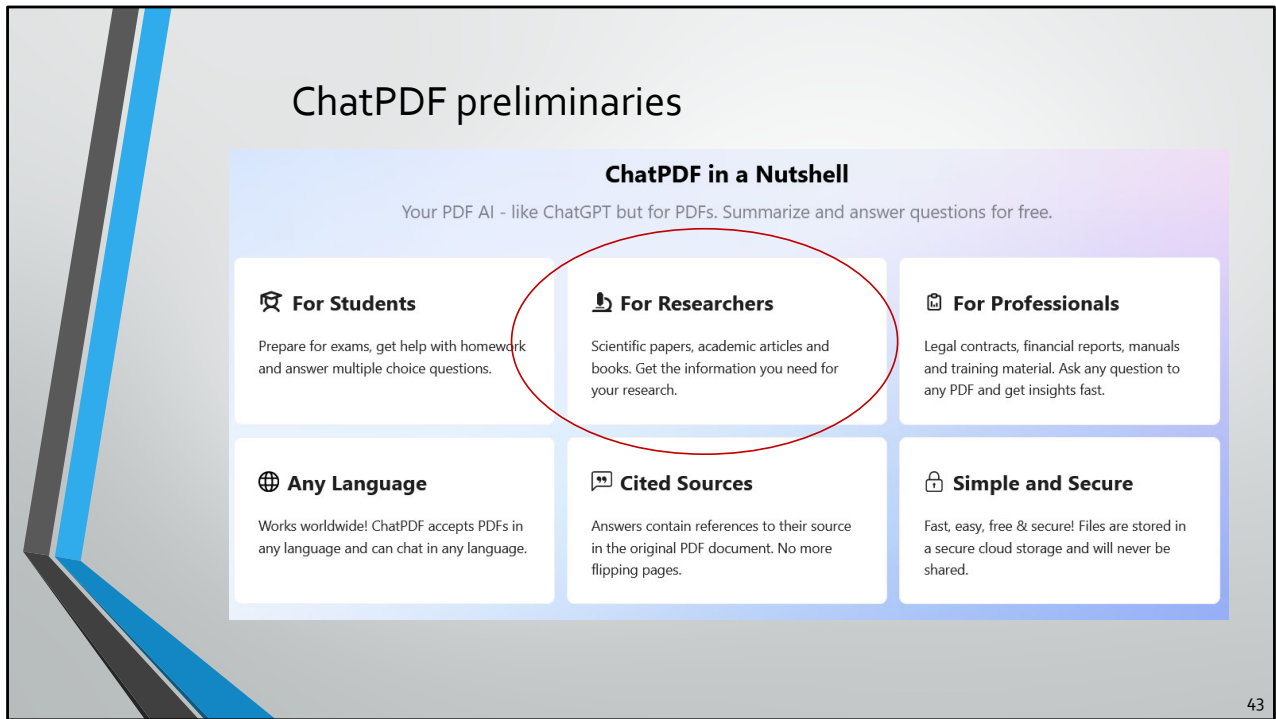
Ok, let's shift gears. Let's talk about how to use AI to summarize articles. As we've seen, it's possible to enter excerpts directly into ChatGPT, but there is a limit to the size of text inputs. The free version of ChatGPT can't access the internet, although it can give the impression it can access URLs. I've added a link to a blog article in the handout discussing this phenomenon. It's actually super interesting. Apparently, ChatGPT is just really good at making stuff up based on the text in the link itself. Similarly, I've read suggestions one can enter a citation into ChatGPT to request a summary of the article, but I'd be worried about this approach because ChatGPT might be more prone to hallucinate with this type of prompt.

Although there are many tools that can summarize articles, I'm going to demonstrate just one tool, ChatPDF, very quickly, because it's popular and easy to use. Before I get to the demonstration, though, I'll discuss some of the issues involved.

## ChatPDF preliminaries



Earlier, I didn't spend any time on the ChatGPT terms of use, but let's take a peek now. On this slide, you can see that the terms state that users "may not use the Services in a way that infringes, misappropriates or violates any person's rights" (<https://openai.com/policies/terms-of-use>), so OpenAI is protected if users submit content to ChatGPT they shouldn't, such as licensed journal articles.



ChatPDF seems to take a very different approach by inviting submissions of many kinds of resources, including articles, as you can see on this snip from their website.

## ChatPDF preliminaries

### ChatPDF in a Nutshell

Your PDF AI - like ChatGPT but for PDFs. Summarize and answer questions for free.

#### For Students

Prepare for exams, get help with homework and answer multiple choice questions.

#### For Researchers

Scientific papers, academic articles and books. Get the information you need for your research.

#### For Professionals

Legal contracts, financial reports, manuals and training material. Ask any question to any PDF and get insights fast.

#### Any Language

Works worldwide! ChatPDF accepts PDFs in any language and can chat in any language.

#### Cited Sources

Answers contain references to their source in the original PDF document. No more flipping pages.

#### Simple and Secure

Fast, easy, free & secure! Files are stored in a secure cloud storage and will never be shared.

... and they promise security. So it seems like there should be no problem submitting journal articles found at the library, right?

## ChatPDF preliminaries

### Are my files secure?

ChatPDF will never share your files with anyone. They are stored on a secure cloud storage and can be deleted at any time.

45

Well, I wasn't so sure, so I tried looking around a little more. First, I checked their FAQ for more information, but the FAQ also makes it seem like any kind of submission would be okay. They even repeat their file security guarantee in plain terms.

## ChatPDF preliminaries

### Are my files secure?

ChatPDF will never share your files with anyone. They are stored on a secure cloud storage and can be deleted at any time.

### 7. PROHIBITED ACTIVITIES

You may not access or use the Services for any purpose other than that for which we make the Services available. The Services may not be used in connection with any commercial endeavors except those that are specifically endorsed or approved by us.

As a user of the Services, you agree not to:

- Upload, share, distribute, or otherwise use any PDF files for which you lack the necessary rights, permissions, or licenses. This encompasses copyrighted material, trade secrets, and confidential documents that you are not explicitly authorized to access or disseminate by the rightful owner(s).

Ultimately, it was only by digging deep into the Terms that I found a prohibition on certain submissions. My concern is that I don't think most users will see or heed these terms.

ChatPDF demonstration

<LIVE>

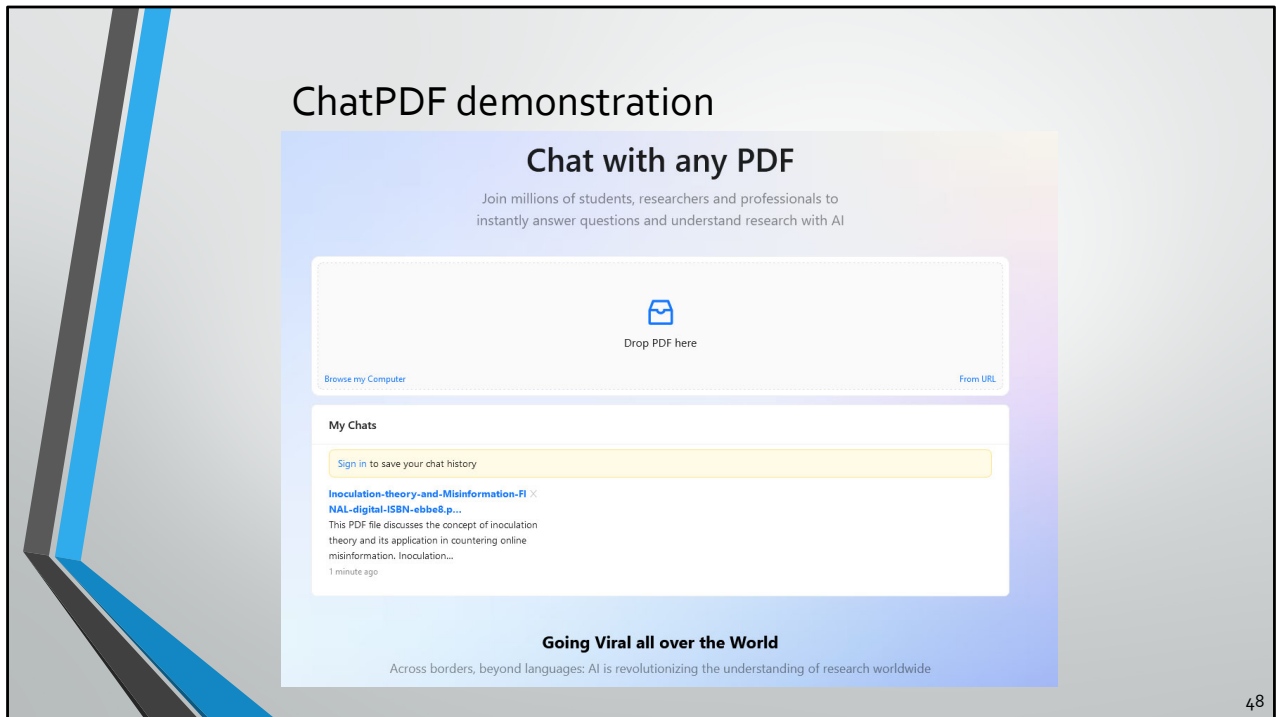
<https://www.chatpdf.com/>

See <https://link.mnsu.edu/aicript>  
for my script.

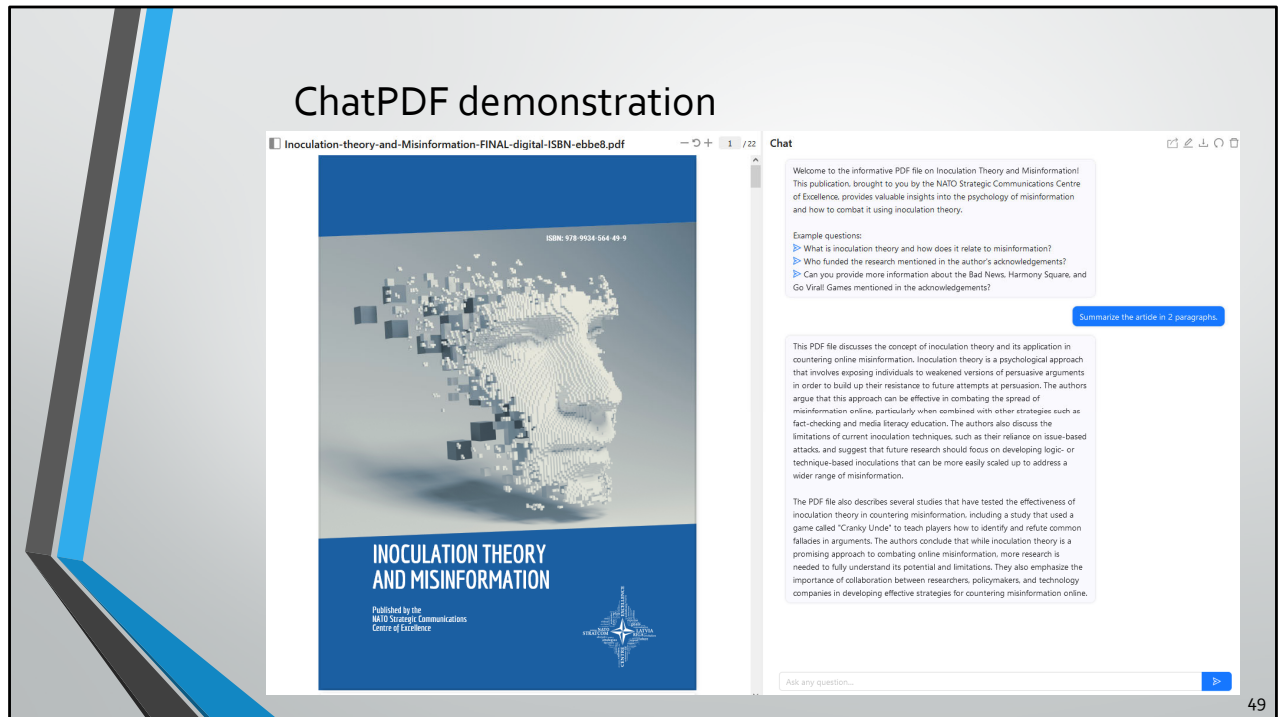
47

But I will today. Fortunately, I have an open access article handy, so let's try that...





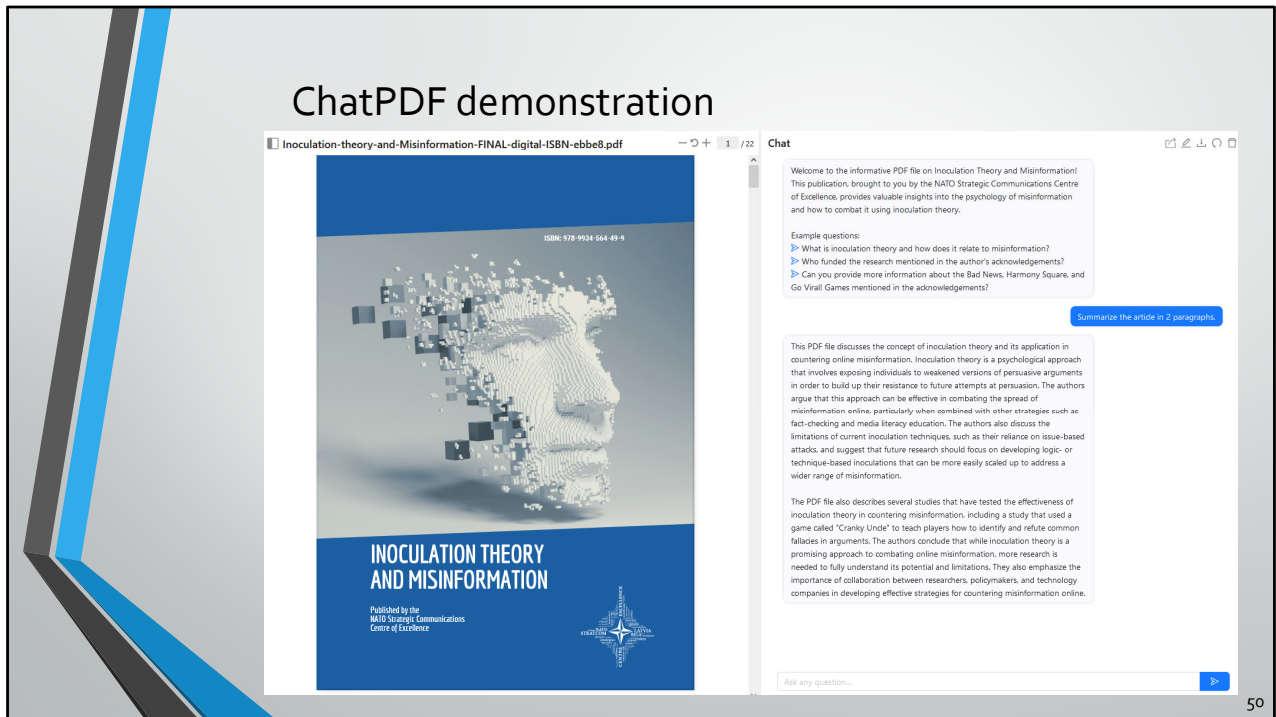
I will begin simply by uploading an open access article I've previously downloaded. The article is entitled "Inoculation Theory and Misinformation," published by the NATO Strategic Communications Centre of Excellence.



ChatPDF will quickly read the article and provide example questions you might ask. For our purposes, let's just ask for a quick 2-paragraph summary.,,,

<Summarize the article in 2 paragraphs.>

As you can see, the results come up quickly, and although we don't have time to read carefully and compare this summary to the article, I can say it's pretty good from previous tests. This is basically just like an abstract, but it's helpful for articles when there is no abstract.



We could ask more pointed questions. For example...

<How does the article define misinformation. Please respond in one sentence.>

---

I won't spend more time on this, but hopefully, this has provided a sense of the possibilities.

## Article summarization issues & discussion

### “User input

When providing user input, the user must take into account, among other things, the ChatGPT terms of use. These state, for example, that the user must not use the service in a way that would harm the rights of third parties. This means that user input must not contain copyrighted data without the permission of the author or rights holder, or that the user must not use confidential data for no reason.

The terms of use also state that if you process personal data when using the service, that you must have made the necessary notifications and obtained the necessary consent and that you certify that the processing is in accordance with applicable laws. The terms of use also provide that the user must contact OpenAI to enter into a processing agreement if the processing is subject to the GDPR or the *California Consumer Privacy Act* (CCPA). For professionals bound by professional secrecy, such as lawyers, it goes without saying that they should not use information subject to professional secrecy as input. The reason for all these restrictions on user input is obvious.

After all, user input is not only used to generate output, but is also reused to improve the service, and thus refine the AI model. The user input ends up on a big pile of data and the user then basically loses control over that data. OpenAI is reportedly working on a professional paid version of ChatGPT, but it is still unclear whether professional user input, which may also contain confidential professional data, will also be used to improve the service. Organisations using the ChatGPT API can, however, request OpenAI not to use their input data to improve the services. But for ordinary users, such an opt-out does not currently seem possible.”

Fiten & Jacobs, 2023: <https://www.timelex.eu/en/blog/chatgpt-and-copyright-many-questions-remain-be-answered>

51

I’m excited by article summarization tools, but I do think librarians need to take the lead to communicate the issues involved. If students or faculty submit articles to tools like ChatPDF, there could be some legal risk for the university, depending on the materials involved. I think it would be fine to submit open access journal articles, but many journals and other materials are covered by both copyright and license restrictions. As we’ve seen, ChatPDF promises secure storage, but so long as your institution or system hasn’t negotiated a license with ChatPDF, or similar services, and we haven’t carved out the right to submit restricted materials to these services in the library’s licenses with publishers, then we really shouldn’t.

## Article summarization issues & discussion

**SPRINGER NATURE** Search EN

Librarians Products Licensing Open Research Tools & Services Blog Contact Stay Informed

### Artificial Intelligence and Machine Learning

Artificial intelligence is one of the most rapidly advancing and controversial topics in scientific research. Escalating computing power, expanding data sets, and algorithms of unprecedented sophistication have led to a massive increase in the number of journal and conference papers referring to AI in recent years. Using AI has enabled efficient research and many examples of its usefulness for academic and industrial development have emerged.

Explore the content and resources on this page and gain insights into how AI & Machine Learning are being explored and applied in scholarly and corporate contexts.

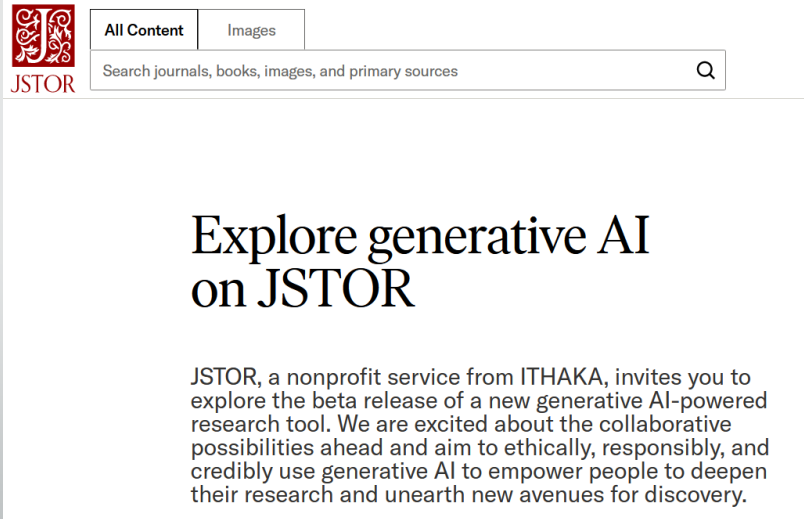
**Contact us today!**  
For information & pricing to add AI and Machine Learning content to your Library

<https://www.springernature.com/gp/librarians/landing/ai-machine-learning>

52

That said, last year, I attended a Springer Nature webinar where they demonstrated how they can summarize articles and create new content from existing articles. It will be interesting to see what happens next, if these kinds of features become more available in our libraries. I think content ownership issues could be a major obstacle to progress. A problem is that our commercial vendors don't work together well – as we know – and the same problems that plague discovery will likely plague AI applications.

## Article summarization issues & discussion



The screenshot shows the JSTOR website interface. At the top left is the JSTOR logo. To its right are two tabs: "All Content" (selected) and "Images". Below these is a search bar with the placeholder text "Search journals, books, images, and primary sources" and a magnifying glass icon. The main content area features the heading "Explore generative AI on JSTOR" in a large, bold font. Below the heading is a paragraph of text: "JSTOR, a nonprofit service from ITHAKA, invites you to explore the beta release of a new generative AI-powered research tool. We are excited about the collaborative possibilities ahead and aim to ethically, responsibly, and credibly use generative AI to empower people to deepen their research and unearth new avenues for discovery."

<https://www.jstor.org/generative-ai-faq>

53

JSTOR is also adding AI augmentation. You can sign up on a waiting list to try this functionality. I've signed up but I haven't been invited to try it yet, so I can't demonstrate it.

## Consensus demonstration

<LIVE>

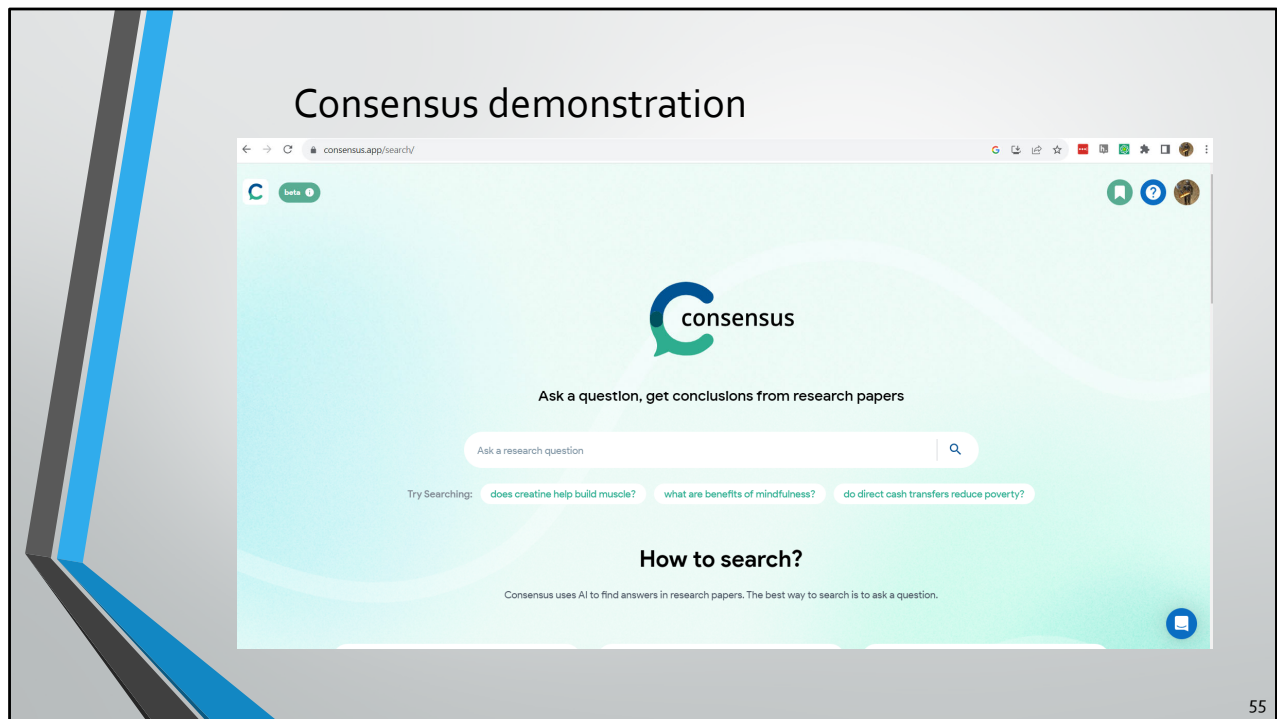
<https://consensus.app/search/>

See <https://link.mnsu.edu/aicript>  
for my script.

54

Let's move on and take a look at another AI tool. Consensus began as a tool to search for journal articles using natural language prompts. When I started using it, it basically just provided short excerpts from articles. Now, Consensus is also integrated with GPT-4 to provide new functionality, including summarization of the 5-10 most appropriate articles addressing the prompt. The new functionality can be tested for free, but there are only a small number of tokens provided per month.

Consensus is both a discovery tool to find journal articles *and* it provides AI summarization, so it might address some of the issues I raised about ChatGPT as an information source. Let's take a look...



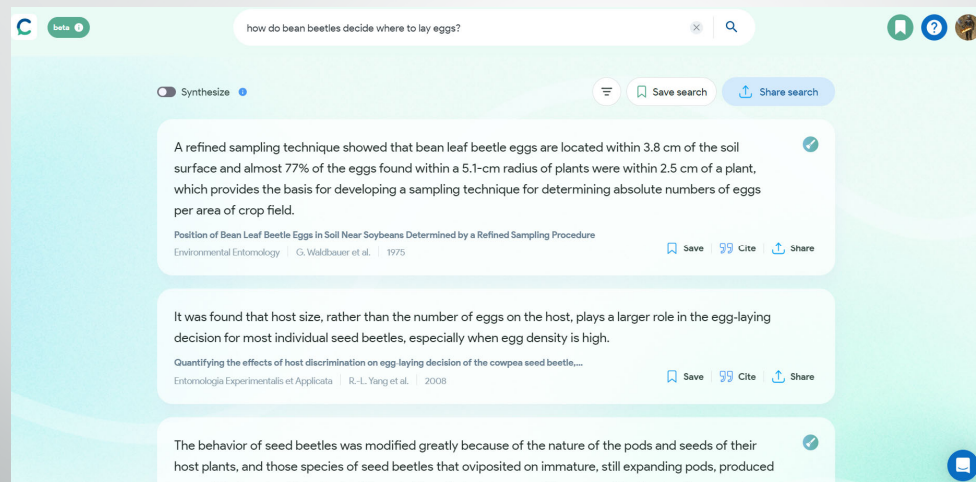
Let's start with the free version. A Consensus subscription costs \$7.99 and students can get a 40% discount.

As a discovery tool, there are some basics to know. So far as I understand it, based on work I did with Consensus in early 2022, I think Consensus was built using the article database developed by Semantic Scholar, which also calls itself an AI-powered research tool. I think Consensus has focused on natural language processing and article summarization, whereas Semantic Scholar has focused on other functionality. I haven't studied either deeply. Based on the little I've done with each, I'd say Consensus has more appeal for undergrads, while Semantic Scholar provides added value for grad students and faculty. According to Semantic Scholar, they have indexed "over 200 million academic papers sourced from publisher partnerships, data providers, and web crawls" (<https://www.semanticscholar.org/about>).

Anyway, my concern today isn't really to dive deeply into these tools, but to make sure you're aware of them and maybe thinking about how to use them. To be honest, I've never spent much time on Consensus because I have expected it to be overtaken by the traditional players in this space. Just a couple of weeks ago, I saw on the Liblicense listserv an announcement that Scopus now integrates some amount of AI support – so I expect we'll see more and more of these kinds of advances.



## Consensus demonstration

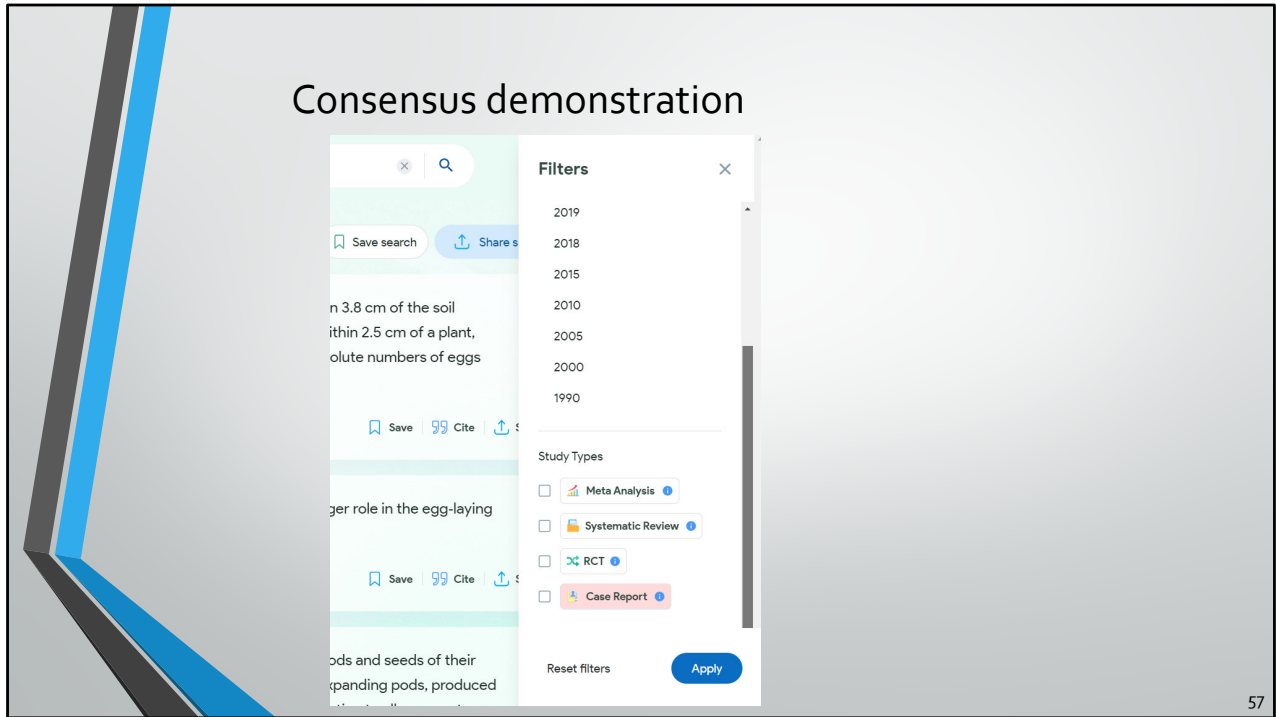


56

Consensus is optimized for input as a question. Here, I will repeat a question I put to ChatGPT. In that case, I said "Provide 3 links to scholarly sources on the oviposition of bean beetles (*Callosobruchus malculatus*).". In this case, I'll ask "How do bean beetles decide where to lay eggs?"

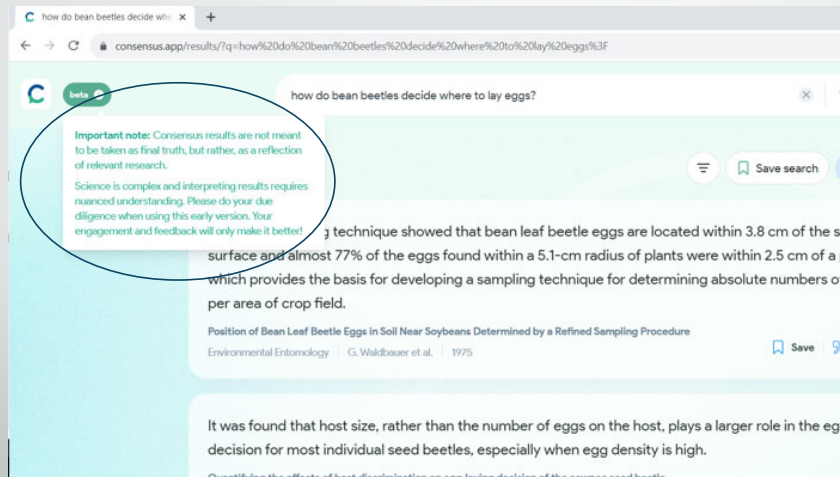
In the result set, you can see that the excerpts directly pertain to the question. I don't know if Consensus would always be quite so successful, but I do think it could be helpful for students. As we all know, students sometimes don't use the best possible discovery interface for a given question. If they use Primo, the results can be very hit or miss, and there's always a ton of junk to slog through. In any case, I do think Consensus, or tools like it, could be worthwhile to add to our toolboxes as librarians, although it can't currently replace our library discovery tools. I don't imagine most students would pay for a subscription or even find this tool on their own, necessarily, but it could be something to use at the reference desk occasionally, or possibly to introduce during instruction sessions for some purposes.

## Consensus demonstration



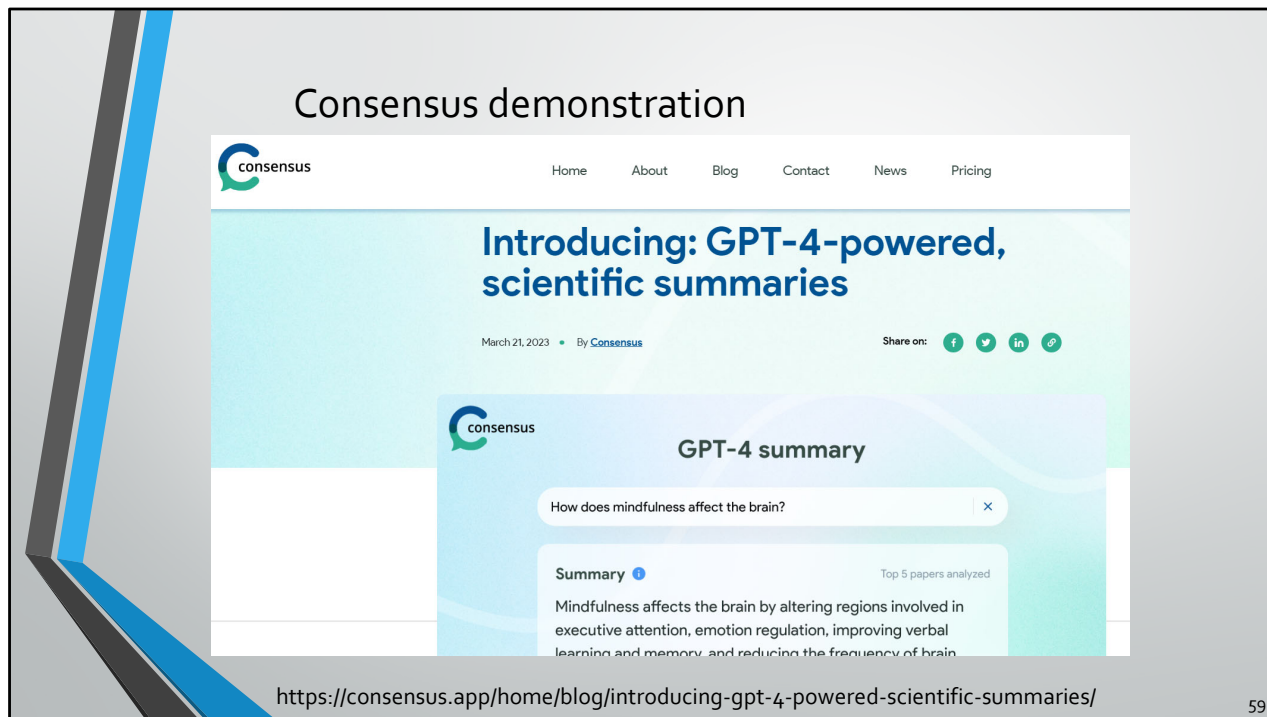
The filters of Consensus are not yet very impressive, but I do like the option to filter by study types.

## Consensus demonstration

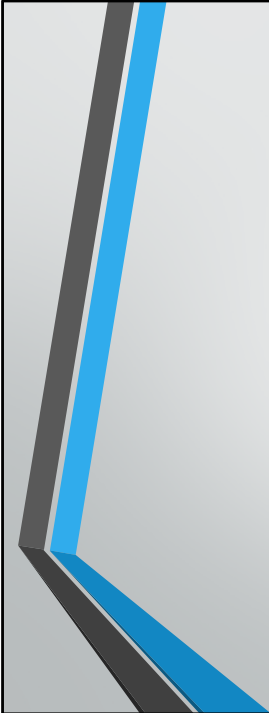


58

Consensus also comes with some caveats, such as this warning about results. But, again, we know from experience that other discovery layers are not perfect either.



What I've shown so far is the basic, free functionality. I haven't used the GPT-4-powered functionality because there are only three free tokens. Let's try it together now. [No screenshot for deck because I saved my tokens for the live demonstration.]



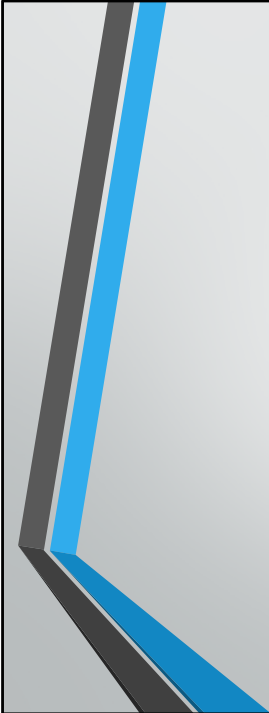
## Additional Thoughts

- Domain Specific Projects

60

So anyway... This has been an introductory session. There are many other topics we could explore... I think, over time, AI discussions and presentations will grow increasingly domain specific. For example, I perform licensing, among other things. I'd like to experiment with a customized ChatBot to support license review. I would not present that experiment to a general audience interested in AI, but to an audience of licensing specialists, possibly through NASIG.

Actually, I can see multiple use cases across the range of my work, including licensing, data analysis, research consultations, and library instruction.



## Additional Thoughts

- Domain Specific Projects
- The Future of Work

61

While planning this session, I initially thought I would provide a section on the future of work, but I realized I would not have enough time. I think this is an important topic, so I've included links in the handout as food for thought. I think AI proficiency will grow increasingly crucial to our students. Actually, I think AI proficiency is an equity issue for the ages. Whether or not we want to teach with AI, I am fully convinced students *\*need\** to gain the advantage of learning with AI, so we should be as pro-active as possible exploring the possibilities.

## Additional Thoughts

Harvard  
Business  
Review

Business And Society | **AI Won't Replace Humans — But Humans With AI Will Replace Humans Without AI**

Business And Society

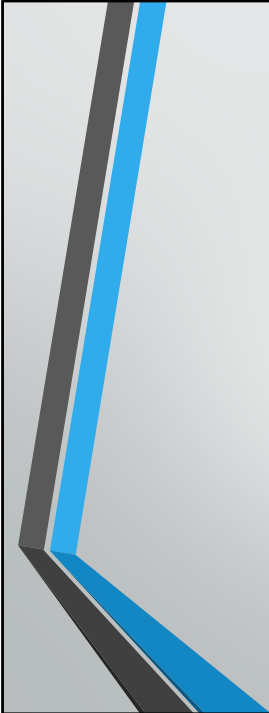
# **AI Won't Replace Humans — But Humans With AI Will Replace Humans Without AI**

August 04, 2023

<https://hbr.org/2023/08/ai-wont-replace-humans-but-humans-with-ai-will-replace-humans-without-ai>

62

In case you haven't seen it, here's the title of a recent HBR article which I think makes the point succinctly.



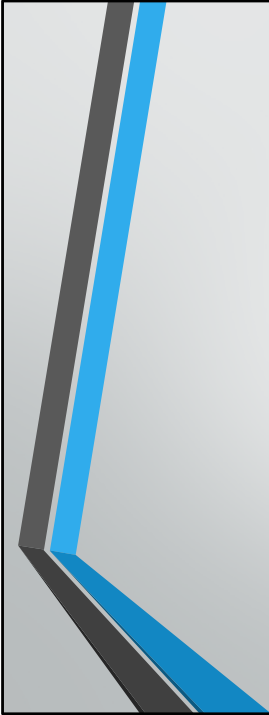
## Additional Thoughts

- Domain Specific Projects
- The Future of Work
- AI & Teaching

63

I recently co-presented with Mark and a colleague from MNSU IT Solutions, Elizabeth Harsma, on “5 Tips for Teaching with AI.” We don’t have time to cover that ground here, but, in case you’re interested, I’ll provide a link to that presentation among the Resources at the end of this deck. I also provided some teaching-related resources in the handout.



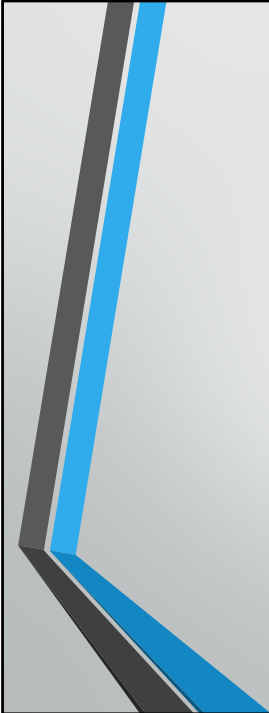


## Additional Thoughts

- Domain Specific Projects
- The Future of Work
- AI & Teaching
- Mark's Session

64

Mark has been thinking about developing a presentation to focus on how AI can support specific library functions, such as reference, instruction, collection development, and open educational resources (OER). If that moves forward, we'll send out more information.



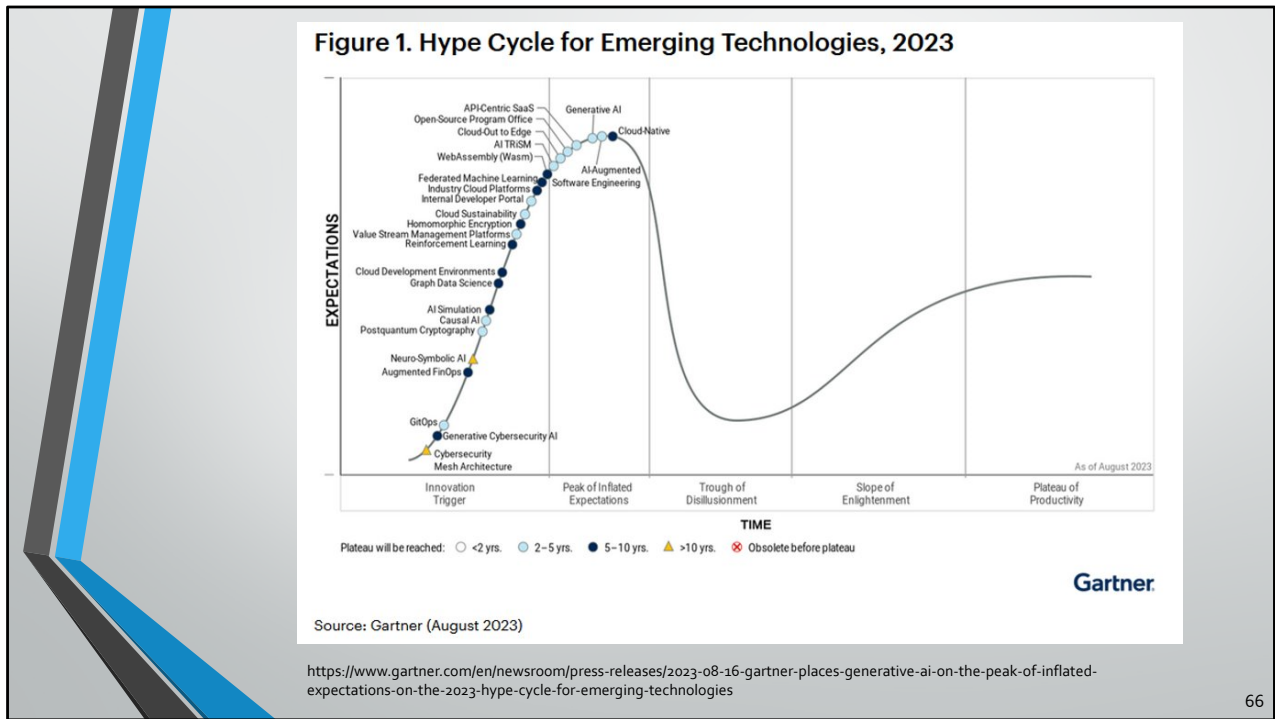
## Additional Thoughts

- Domain Specific Projects
- The Future of Work
- AI & Teaching
- Mark's Session
  
- But isn't this the end of the world???!

65

There are numerous other topics we could cover, including the shadow side of AI, but my goal was to provide an introductory session to encourage regional librarians to engage with AI because librarians should be providing leadership on their campuses, including, but not limited to information literacy, copyright, and licensing concerns. The best quick overview of AI's shadow I've seen was provided by a librarian from the Alder Graduate School of Education by the name of Ray Pun. He was a panelist for the Sage Publishing Critical Thinking Bootcamp on August 8, 2023. I don't think a recording has been posted yet, but it would be worthwhile to look out for it. I'm sure many other librarians will also raise concerns over time.

For what it's worth, I read recently that Claude 2, Anthropic's alternative to ChatGPT is the most ethical option among the bunch, but you'll need to investigate that for yourself.



Finally, I think I should mention that Gartner is already predicting an end to the AI hype cycle. They expect a “trough of disillusionment” coming soon, but this just means (to me) that the real work to develop meaningful applications within our areas of expertise is just starting...

## Resources

<https://link.mnsu.edu/aideck>

<https://link.mnsu.edu/aascript>

<https://link.mnsu.edu/airesources>

<https://link.mnsu.edu/aislides>

67

Once again, please let me remind you that you can access this presentation and the script for my demonstrations at the links above. At the bottom of the screen is the link to the presentation '5 Tips for Teaching with AI' I previously mentioned. I have also prepared a handout available at [link.mnsu.edu/airesources](https://link.mnsu.edu/airesources). Let's take a quick look at the handout... (if there's time).

## Resources

3. Additional Resources
  - a. Generative AI & Teaching
    - i. [AI 101 for Teachers](#) (links to a Code.org course)
    - ii. [Teaching and Generative Artificial Intelligence](#) (links to MNSU Center for Excellence in Teaching & Learning (CETL) website)
    - iii. [A Teacher's Prompt Guide to ChatGPT](#) (links to a PDF posted on a blog entitled "User Generated Education" by Jackie Gerstein, Ed.D.)
    - iv. [100+ Creative ideas to use AI in education](#) (links to Google slides linked to the #creativeHE website, March, 2023. #creativeHE is a group who "share a common interest in creativity and innovation in learning, teaching and research.")
  - b. The Future of Work
    - i. [Generative AI and the future of work in America](#) (links to McKinsey Global Institute report, 7/26/2023)
    - ii. [Which U.S. Workers Are More Exposed to AI on Their Jobs?](#) (links to Pew Research Center report, 7/26/2023)
    - iii. [AI Won't Replace Humans — But Humans With AI Will Replace Humans Without AI](#) (links to Harvard Business Review article, 8/4/2023)
    - iv. [Gartner Places Generative AI on the Peak of Inflated Expectations on the 2023 Hype Cycle for Emerging Technologies](#) (links to Gartner article, 8/16/2023)
  - c. More on Prompts
    - i. [ChatGPT Prompt Engineering for Developers](#) (This is a 1-hour webinar, not publicly listed)

<https://link.mnsu.edu/airesources>

[No script.]

## Resources

- e. Data Analysis
  - i. [Data Analysis For All: Step-by-Step Guide To Do Sentiment & Emotions Analysis of Tweets Using ChatGPT Code Interpreter](#) (links to AI Tools Club article, 7/11/2023)
  - ii. [ChatGPT-Powered Data Exploration: Unlock Hidden Insights in Your Dataset](#) (links to KDnuggets article, 7/17/2023)
  - iii. [Integrating GenAI into "Thinking Like a Data Scientist" Methodology](#) (links to a Data Science Central article, 8/6/2023 – this is the first of a series of 2, so far)
- f. Fun stuff (?)
  - i. The University of Cambridge created [a misinformation susceptibility](#) test using ChatGPT. (links to University of Cambridge article)
  - ii. Watch a 12-minute film with images entirely generated by AI: [The Frost](#). (links to MIT Technology Review article, 6/1/2023)
  - iii. [Aliens: The Musical](#) (sans music). (links to [digitaltrends](#) article, 5/1/2023)
  - iv. [AI Dungeon](#) is an AI-generated text adventure game. (links to AI Dungeon website)
  - v. [With "thanabots," ChatGPT is making it possible to talk to the dead](#) (links to Big Think article, 8/4/2023)
  - vi. [Can Robots Be Beatniks? An AI-Powered Allen Ginsberg Finds Out](#) (links to Decrypt Scene article, 8/20/2023)
  - vii. [This AI Paper Introduces A Comprehensive RDF Dataset With Over 26 Billion Triples Covering Scholarly Data Across All Scientific Disciplines](#) (links to [Marktechpost](#) article, 8/21/2023; not exactly fun, but probably interesting for librarians...)

<https://link.mnsu.edu/airesources>

[No script.]

## References

Ali, F. (2023, April 11). *GPT-1 to GPT-4: Each of OpenAI's GPT models explained and compared*. Make Use Of. <https://www.makeuseof.com/gpt-models-explained-and-compared/>

Cheng, M. (2023, July 10). "Shadow libraries" are at the heart of the mounting copyright lawsuits against OpenAI. Quartz. <https://qz.com/shadow-libraries-are-at-the-heart-of-the-mounting-cop-1850621671>

Cooper, K. (2021, November 1). *OpenAI GPT-3: Everything you need to know*. Springboard. <https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/>

Feng, S., Park, C.Y., Liu, Y. & Tsvetkov, Y. (2023, July). From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 1, 11737–11762. <https://aclanthology.org/2023.acl-long.656.pdf>

Fiten, B. & Jacobs, E. (2023, March 23). *ChatGPT and copyright: many questions remain to be answered*. Timelex. <https://www.timelex.eu/en/blog/chatgpt-and-copyright-many-questions-remain-be-answered>

Here is a list of references in the slides.

## References, cont'd

Fulford, I. & Ng, A. (n.d.). *ChatGPT prompt engineering for developers*. DeepLearning.ai. <https://learn.deeplearning.ai/chatgpt-prompt-eng/lesson/7/expanding>

Gartner. (2023, August 16). Gartner Places Generative AI on the Peak of Inflated Expectations on the 2023 Hype Cycle for Emerging Technologies. Gartner. <https://www.gartner.com/en/newsroom/press-releases/2023-08-16-gartner-places-generative-ai-on-the-peak-of-inflated-expectations-on-the-2023-hype-cycle-for-emerging-technologies>

Helms, S. & Krieser, J. (2023, March). *Copyright chaos: Legal implications of generative AI*. <https://www.bloomberglaw.com/external/document/XDDQ1PNK000000/copyrights-professional-perspective-copyright-chaos-legal-implic>

Henshall, W. (2023, August 2). *4 charts that show why AI progress is unlikely to slow down*. Time. <https://time.com/6300942/ai-progress-charts/>

Myers, A. (2023, June 5). *Reexamining "Fair Use" in the age of AI*. Stanford University Human-Centered Artificial Intelligence (HAI). <https://hai.stanford.edu/news/reexamining-fair-use-age-ai>



## References, cont'd

OpenAI. (2023, March 14). *GPT-4*. OpenAI. <https://openai.com/research/gpt-4>

Roberts, G. (n.d.). *AI training datasets: the Books1+Books2 that big AI eats for breakfast*. Musings of Freedom. <https://gregoreite.com/drilling-down-details-on-the-ai-training-datasets/>

P.S.

Gamma

Templates Login Sign up for free

Never start from scratch

## Generate docs, decks & webpages in seconds

Create a working **presentation**, **document** or **webpage** you can refine and customize in under a minute, using our powerful **AI generator**.

Gamma allows me to package up information in **ways I can't with slides**, while still creating good flow for my presentations.

Ann Marie, Director of Product at Koalaft

<https://gamma.app/?ref=aitoolsclub.com&lng=en>

<https://link.mnsu.edu/aifeedback>

By the way, we could also have a long talk about AI tools for office productivity. This slide shows Gamma, an AI that generates presentations. I created this presentation the old-fashioned way, but maybe I should have asked Gamma? Please let me know what you think. To provide feedback, go to [link.mnsu.edu/aifeedback](https://link.mnsu.edu/aifeedback).



*Thank you!*

Nat Gustafson-Sundell  
gustan2@mnsu.edu

MarkMcCullough  
mark.mccullough@mnsu.edu