AI in Academic Work: Capabilities, Limitations, and Responsible Use

May 11th, 2026

By Lois Wong, AI Librarian

THE UNIVERSITY OF CHICAGO | THE UNIVERSITY OF CHICAGO Library

Agenda

1. Introduction
2. How Language Models Work
3. Hallucination and Bias
4. Use Cases in Academic Work
5. Thinking about Responsible Use

My Background

Education

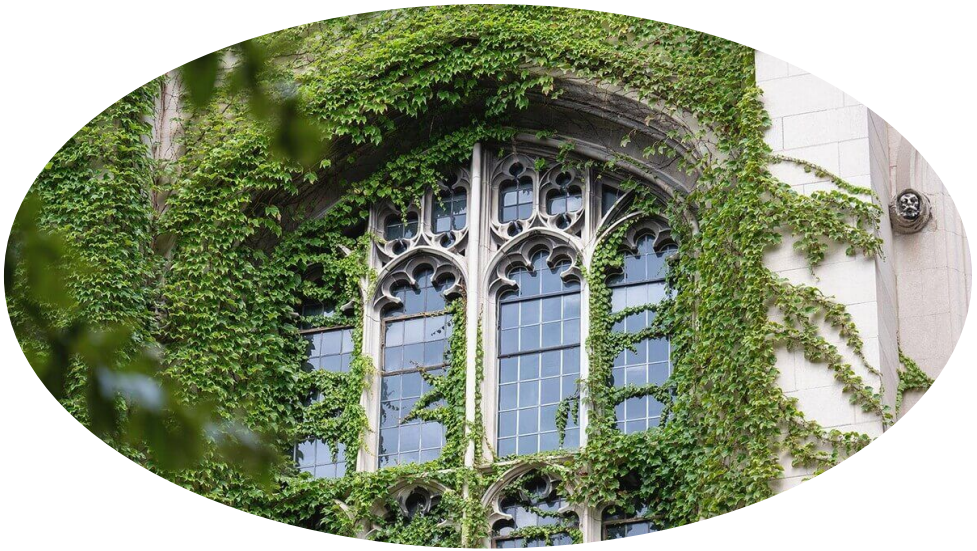
- **BA in Linguistics**, UC Berkeley
- **MSE in Computer Science**, Johns Hopkins University

Professional Background

- AI Education Teaching Assistant, **Apple**
- Data Analytics & Systems Intern, **International Rescue Committee**
- Founding DevRel Engineer, **Novita AI**
- AI Librarian since September 2025

What is AI

1. **Aligning on Key Terms**
2. **How LMs work - through the lens of hallucination**



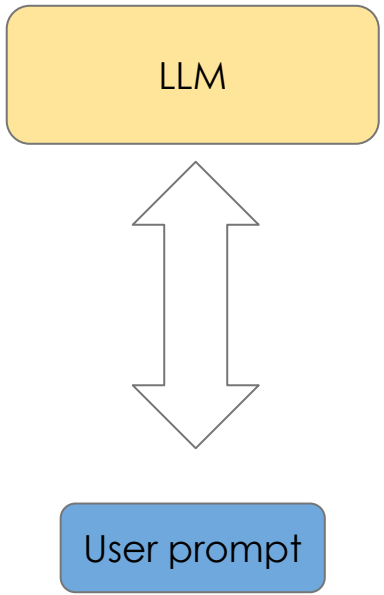
Terminology

- **AI -** The idea/science/ engineering of simulating human intelligence
- **AI or ML Model -** general term for computer programs that **learn from data** to make predictions, generate content, or classify information
- **LLMs/LMs/SLMs -** takes an input **in natural language** (prompt) and produces a relevant output (response) also in natural language
    - Parameter count: LLM - billions ; SLM - millions/hundreds of millions
    - This is cool because it's like talking to a person

Basic Terminology

- **AI applications / tools / assistants -** specific products that usually has a model in their architecture and have a user-friendly interface
- **Inference:** Using an LM to generate outputs (predictions, insights, etc.)
    - Ranges from simple user interactions (e.g., chatbots) to a layer within larger workflows
- **LM APIs** (Application Programming Interface) **-** what enables you to connect your app/project to a model)

Interacting with the model via API

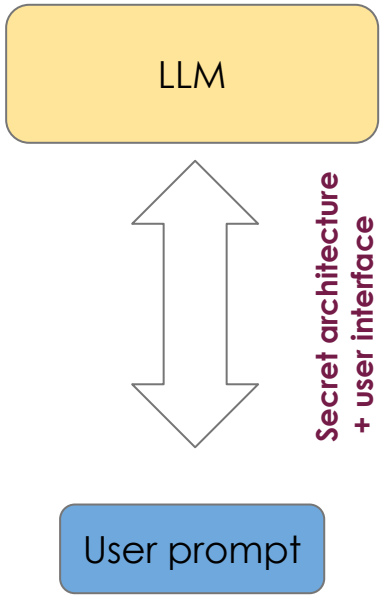


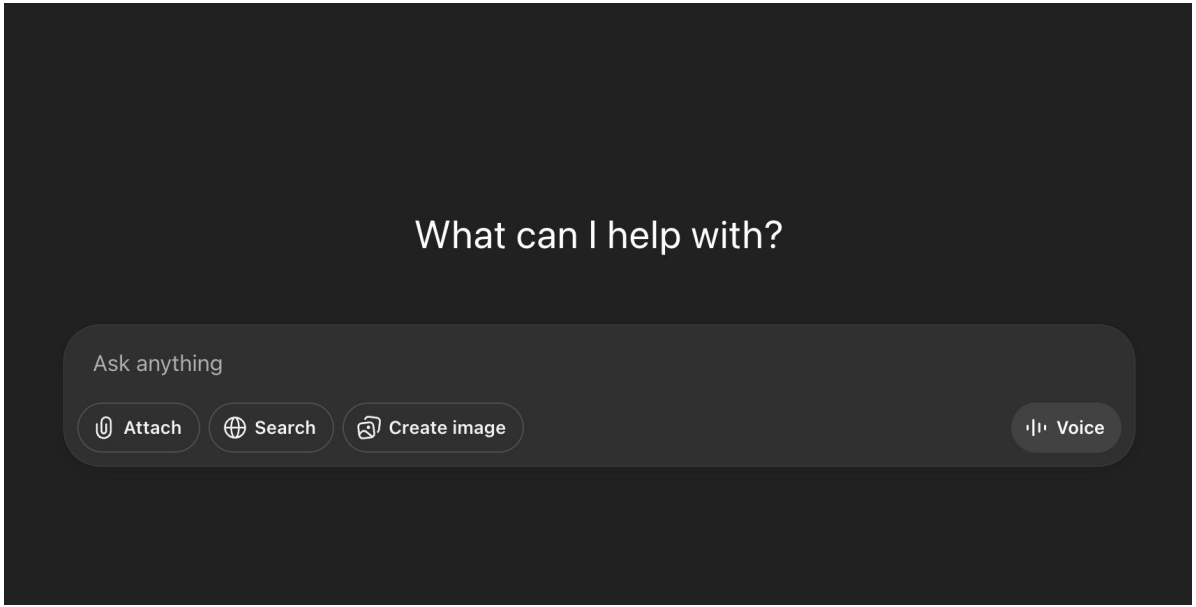

```python
import openai
from openai import OpenAI
import numpy as np

client = OpenAI(
  api_key= "secret"
)

response = client.chat.completions.create(
        model="gpt-4o-mini",
        messages=[
            {"role": "system", "content": "You are a helpful assistant"},
            {"role": "user", "content": "what's up"}
        ],
        #max_tokens=1000,
    )
chatgpt_response = response.choices[0].message.content
print(chatgpt_response)
```

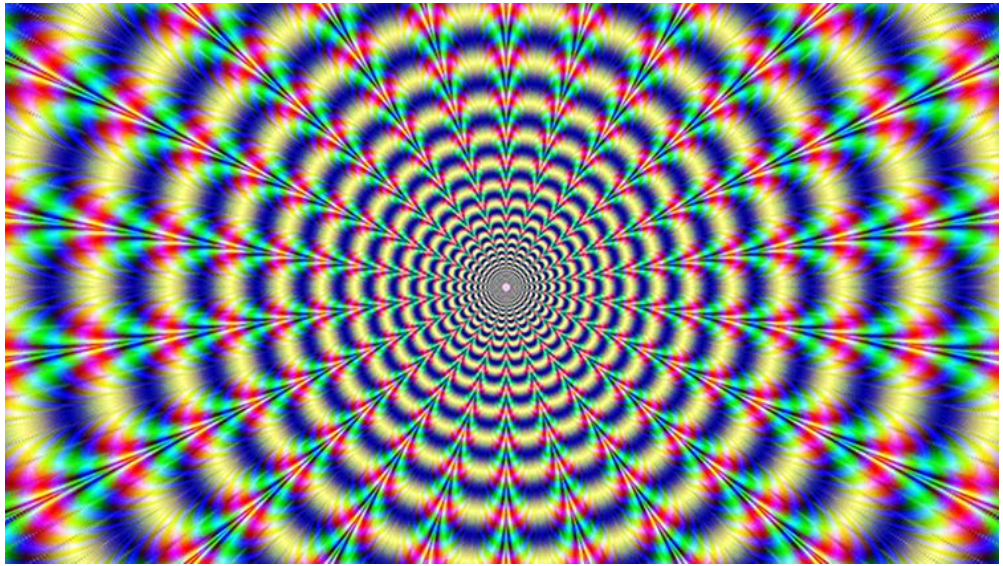
Interacting with the model via OpenAI's User Interface





How LMs work - through the lens of Hallucination

- AI cannot reliably output ‘true’ information
- Prone to **hallucination**, a catch-all phrase for responding with inaccurate or false information



Hallucination

Example: Lit Review on the role of nature in Macbeth

- **Rosenberg, Marvin. "The Masks of Macbeth."** In this influential work, Rosenberg explores the psychological and symbolic dimensions of Macbeth. He delves into the concept of masks worn by the characters, emphasizing how they reveal and conceal elements of human nature. Rosenberg's analysis provides valuable insights into the complex interplay between the internal and external aspects of the characters, shedding light on the role of nature as a reflection of their inner struggles.
- **Nevo, Ruth. "Forms of Expansion: Recent Studies in Shakespeare."** Nevo's critical examination encompasses various aspects of Shakespearean works, including Macbeth. Her exploration of the play delves into the thematic significance of nature and its impact on character development. Nevo's work serves to contextualize the broader trends in Shakespearean scholarship, offering a lens through which to understand the evolving interpretations of the natural world in Macbeth.
- **Foakes, R. A. "Nature in Shakespearean Tragedy."** Foakes provides a broader perspective by examining the thematic treatment of nature across several Shakespearean tragedies. In relation to Macbeth, Foakes discusses the broader context of nature as a recurring motif in Shakespeare's works and its unique manifestation in Macbeth. This comparative approach contributes to a more comprehensive understanding of the thematic nuances specific to Macbeth.
- **Orgel, Stephen. "Macbeth and the Imperial Theme."** Orgel's analysis focuses on the imperial theme within Macbeth, exploring the connection between political ambition and the disruption of the natural order. His work sheds light on the broader societal implications of the characters' actions and ambitions, emphasizing how the natural world serves as a reflection of the political and social dynamics within the play.

Evaluate its output: which of these sources exist?

Hallucination

Example: Lit Review on the role of nature in Macbeth

- **Rosenberg, Marvin. "The Masks of Macbeth."** In this influential work, Rosenberg explores the psychological and symbolic dimensions of Macbeth. He delves into the concept of masks worn by the characters, emphasizing how they reveal and conceal elements of human nature. Rosenberg's analysis provides valuable insights into the complex interplay between the internal and external aspects of the characters, shedding light on the role of nature as a reflection of their inner struggles.
- **Nevo, Ruth. "Forms of Expansion: Recent Studies in Shakespeare."** Nevo's critical examination encompasses various aspects of Shakespearean works, including Macbeth. Her exploration of the play delves into the thematic significance of nature and its impact on character development. Nevo's work serves to contextualize the broader trends in Shakespearean scholarship, offering a lens through which to understand the evolving interpretations of the natural world in Macbeth.
- **Foakes, R. A. "Nature in Shakespearean Tragedy."** Foakes provides a broader perspective by examining the thematic treatment of nature across several Shakespearean tragedies. In relation to Macbeth, Foakes discusses the broader context of nature as a recurring motif in Shakespeare's works and its unique manifestation in Macbeth. This comparative approach contributes to a more comprehensive understanding of the thematic nuances specific to Macbeth.
- **Orgel, Stephen. "Macbeth and the Imperial Theme."** Orgel's analysis focuses on the imperial theme within Macbeth, exploring the connection between political ambition and the disruption of the natural order. His work sheds light on the broader societal implications of the characters' actions and ambitions, emphasizing how the natural world serves as a reflection of the political and social dynamics within the play.

Only this one is real!

Hallucination

Shakespearean scholar, but no article by this name exists

- **Nevo, Ruth. "Forms of Expansion: Recent Studies in Shakespeare."** Nevo's critical examination encompasses various aspects of Shakespearean works, including Macbeth. Her exploration of the play delves into the thematic significance of nature and its impact on character development. Nevo's work serves to contextualize the broader trends in Shakespearean scholarship, offering a lens through which to understand the evolving interpretations of the natural world in Macbeth.

Shakespearean scholar and real article, but not by this author

- **Foakes, R. A. "Nature in Shakespearean Tragedy."** Foakes provides a broader perspective by examining the thematic treatment of nature across several Shakespearean tragedies. In relation to Macbeth, Foakes discusses the broader context of nature as a recurring motif in Shakespeare's works and its unique manifestation in Macbeth. This comparative approach contributes to a more comprehensive understanding of the thematic nuances specific to Macbeth. 1, 2, 3, 4

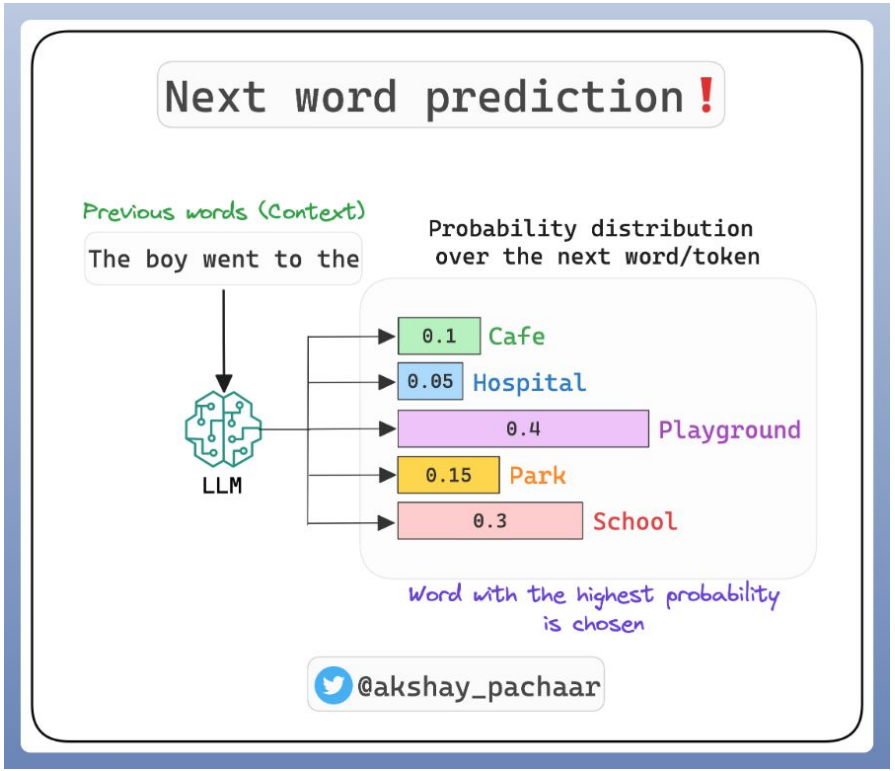
Shakespearean scholar but this article doesn't exist

- **Orgel, Stephen. "Macbeth and the Imperial Theme."** Orgel's analysis focuses on the imperial theme within Macbeth, exploring the connection between political ambition and the disruption of the natural order. His work sheds light on the broader societal implications of the characters' actions and ambitions, emphasizing how the natural world serves as a reflection of the political and social dynamics within the play.

https://chat.openai.com/share/9b0706f2-2d16-45d1-8707-77356fbe1323

Why Do LMs Hallucinate?

- Fundamentally, LM outputs are determined by predicting the next word in a sequence
- It doesn't "know" facts, it only generates what it thinks is most likely based on its training data and context
- This explains why hallucinations often contain partial truths



Working with Hallucination

Theory: It's a data problem

- Training data are not always trustworthy (e.g. Reddit posts)
- Relevant data to answer the question may not exist in the model's knowledge

If you were forced to read the entire internet once and then asked detailed questions later, you'd probably be hallucinating too, especially if you couldn't look anything up.

Grounding the Model

Grounding: providing the LM with external information sources instead of having it rely solely on knowledge it learned from training

1. **RAG** - providing the model with external data sources to draw from
2. **In-context learning** - giving the model helpful stuff in the prompt
    - e.g. including a relevant PDF into your prompt

Hallucination vs Bias vs Error

- **Error**: incorrect output
- **Hallucination**: a type of error specific to GenAI and its probabilistic nature
  - Associated with confidence, sounding believable, and being ungrounded
- **Bias**: outputs reflecting patterns learned from data
  - Models reflect patterns in data
  - These patterns include systemic associations
  - Man : Programmer :: Woman : Homemaker

Measuring Bias

- Bolukbasi et al. finds gender bias in word embeddings in “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings” (2016)
- **Word embeddings** represent words numerically (points in a vector space, where distance reflects similarity in meaning)
  - Enables operations like ‘king – man + woman ≈ queen’
  - *Similarity* is based on the **Distributional Hypothesis**
  - “You shall know a word by the company it keeps” – John Firth
  - Idea: similar words exist in similar contexts

Word Embeddings

- Paper: authors remove gender pair association for gender neutral words

Gender stereotype *she-he* analogies

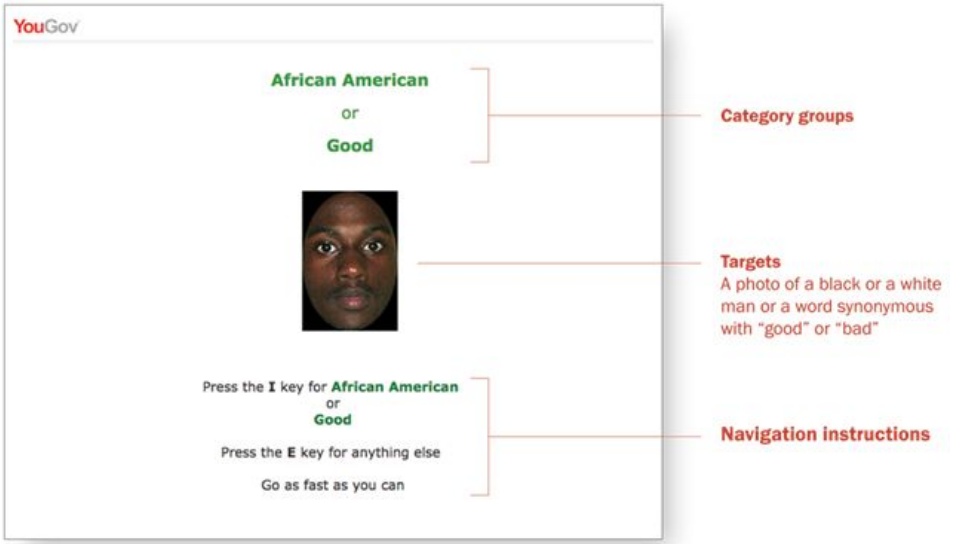
sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant

Gender appropriate *she-he* analogies

queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

https://proceedings.neurips.cc/paper_files/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf

Does Bias Count as Model Error?



Implicit Association Test (psychology)

- Reveals subconscious associations
- Measures how quickly people associate concepts (e.g., “male” or “female”) with attributes (e.g., “career” or “family”), with faster pairings indicating stronger implicit bias

https://www.pewresearch.org/social-trends/2015/08/19/appendix-a-methodology-3/

Does Bias Count as Model Error?

Studies show that the bias in LMs reflect bias in humans

- Female names more associated with family words than career words (Nosek et al., 2002)
- Candidates with White names 50% more likely to be offered an interview than those with Black names after sending 5,000 identical resumes to 1,300 jobs (Bertrand and Mullainathan, 2004)
- Recommended paper: “Semantics derived automatically from language corpora necessarily contain human biases” (Caliskan et al., 2016)

Food for Thought

- What is a good model?
  - Accurate representation of a system
  - Ideal representation of a system
  - Something else
- “AI is a mirror of ourselves, not as we ought to be or could be, but as we already are and have long been.” – Shannon Vallor

Resulting Research

Research emerged from this new modality of quantifying/measuring bias

- “Word embeddings quantify 100 years of gender and ethnic stereotypes” (Garg et al.)
- “Using word embeddings to investigate cultural biases” (Durrheim et al.)

Word Embeddings

The top 10 occupations most closely associated with each ethnic group in the Google News embedding

Hispanic	Asian	White
Housekeeper	Professor	Smith
Mason	Official	Blacksmith
Artist	Secretary	Surveyor
Janitor	Conductor	Sheriff
Dancer	Physicist	Weaver
Mechanic	Scientist	Administrator
Photographer	Chemist	Mason
Baker	Tailor	Statistician
Cashier	Accountant	Clergy
Driver	Engineer	Photographer

https://www.pnas.org/doi/10.1073/pnas.1720347115

AI in Academic Work

1. Research & Discovery
    - Search / Retrieval
    - RAG and MCP
2. Reading & Synthesis
    - Chunking and handling lots of text
3. Writing support
    - Prompt engineering
4. Critique and adversarial testing
    - System vs user prompts

Search and Discovery

AI-enabled search and AI tools like Claude allows us to search more intuitively by asking questions in natural language instead of carefully constructing queries

- Traditional search requires knowing the right terms

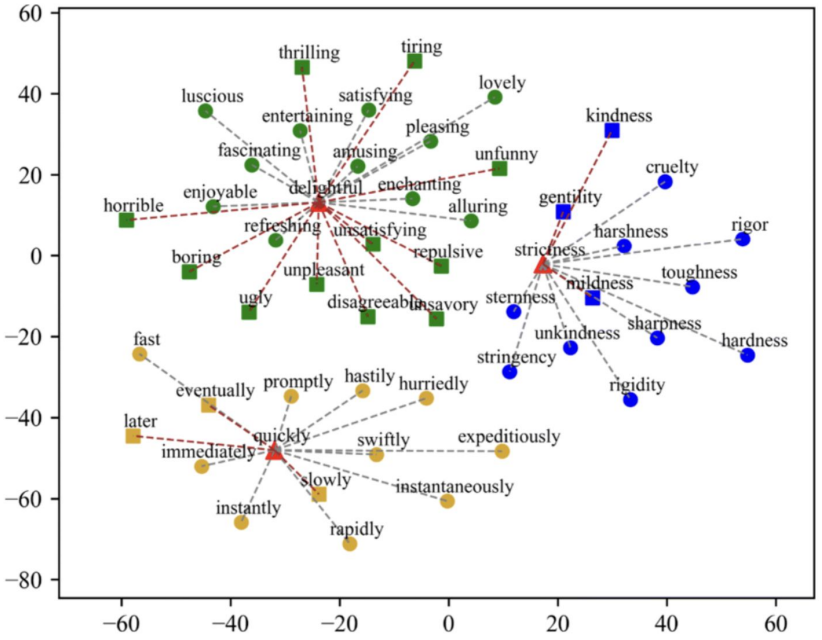
Traditional Search Query: "algorithmic bias" AND "criminal justice" NOT "facial recognition"

AI-Assisted Search Query: “How does algorithmic bias affect work in criminal justice? Give me examples outside facial recognition”

Search and Discovery

Key Disruptions

- Shift from keyword -> natural language search
- Now, search is based on learned semantic representations rather than manually constructed keywords or ontologies
- We can measure the similarity of words (vector distance)



https://www.researchgate.net/figure/Visualization-of-the-word-similarity-relationships_fig16_358362426

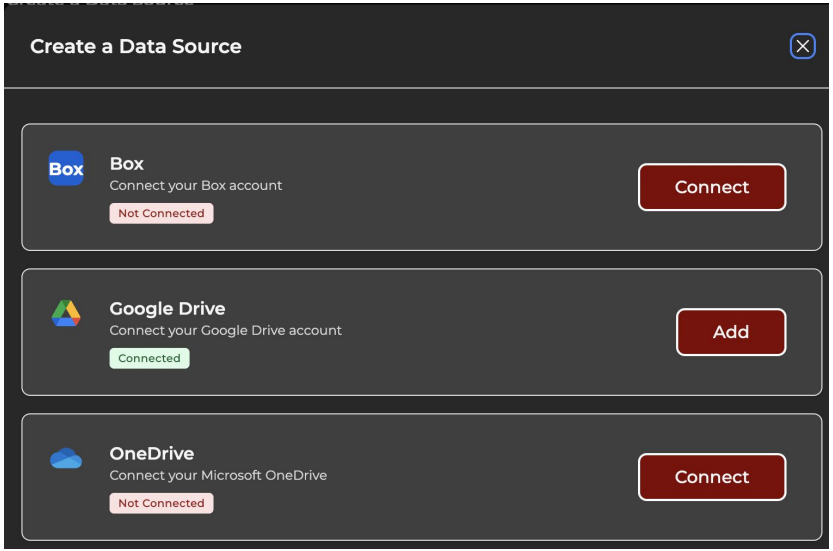
Searching Within Papers

Phoenix AI allows us to connect data sources (a form of RAG)

- Ex: Connect a Google Drive folder of papers and ask specific questions about those papers

Example Use Cases

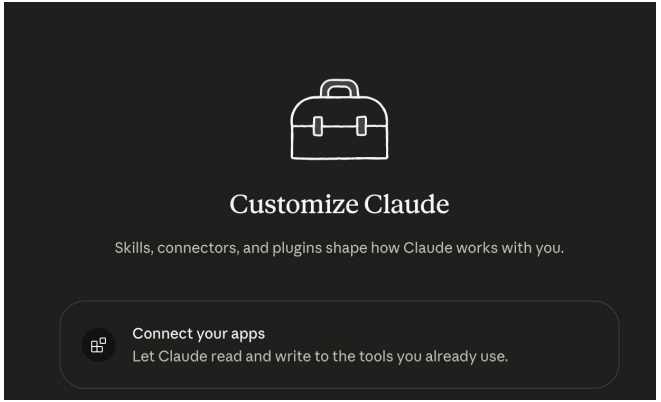
- Ask which of the papers discuss a certain topic
- Synthesize the papers into a lit review



Searching Within Papers

MCP (Model Context Protocol): standardized way for AI systems to connect to external data sources, tools, and systems

- **Wiley Scholar Gateway** - MCP server that gives Claude (client) access to Wiley's papers
- Enables Claude to ground chatbot outputs in Wiley sources
- Improves accuracy and transparency
- Conceptually, it functions like RAG, except the owner of the external data source controls and maintains it
- Unlike LLMs which have a knowledge cutoff date, MCP supports dynamic/live data access



RAG vs MCP (POV Developer)

RAG Workflow	MCP Workflow
Collect documents (scraping)	No need to build or manage your own database
Split into chunks	Data is already structured and maintained by the source
Compute embeddings and store in a vectorDB	
Maintain and update the database (add new data files)	

Chunking

Chunking enables AI to read, compare, and synthesize large amounts of text

- Breaking large texts (documents, books) into smaller, manageable pieces
- Performance often drops when you give the model too much text at once
- Models can lose structure, miss key points, or hallucinate more

Common use cases

- Summarizing long documents
- Comparing and contrasting two or more papers
- Synthesizing insights across many sources / writing a lit review

Chunking

Using AI effectively when dealing with a lot of text

- When using AI to proofread a document, give it one paragraph at a time
- Instead of asking for summaries of 10 files, summarize each individually and then ask for a synthesis
- Trial and error to find the right chunk size - optimize between efficiency and performance
- Note: Using data connectors on Phoenix AI has chunking built into the backend

Using AI in Writing

Use Cases

- Start with one piece of content and ask AI to adapt it for different audiences and formats
- Research → accessible summary / abstract
- Email draft → emails for different audiences and levels of formality
- Long document → executive summary
- Email announcement → LinkedIn post → X post (good for marketing)

Prompt Engineering

Terms that people like to use (but refer to very intuitive concepts)

- **In context learning** - provide examples within the prompt so the model picks up the pattern
  - *Example emails to demonstrate writing tone/style*
  - *Prior meeting minutes or other documents with a defined format/structure you want to replicate*

Zero/one/few-shot learning - give it zero/one/a few examples of the desired output

  - Subset of in-context learning

Prompt Engineering

Terms that people like to use (but refer to very intuitive concepts)

- **Personas** - tell the model who to be to shape tone and priorities
  - *You are a marketing professional with 10 years of experience in consumer tech.*

- **Chain-of-thought** - ask it to show its work (like math problems)
  - *Solve this problem step by step and explain your reasoning: If a train travels 60 miles per hour for 2.5 hours, how far does it go?*

Using AI for Critique and Adversarial Testing

- Models can be overly agreeable or sycophantic
- Motivation: user retention, built to be an agreeable “helpful assistant”

Responses depend on

- System prompts: default instructions guiding behavior
- User prompts: your input

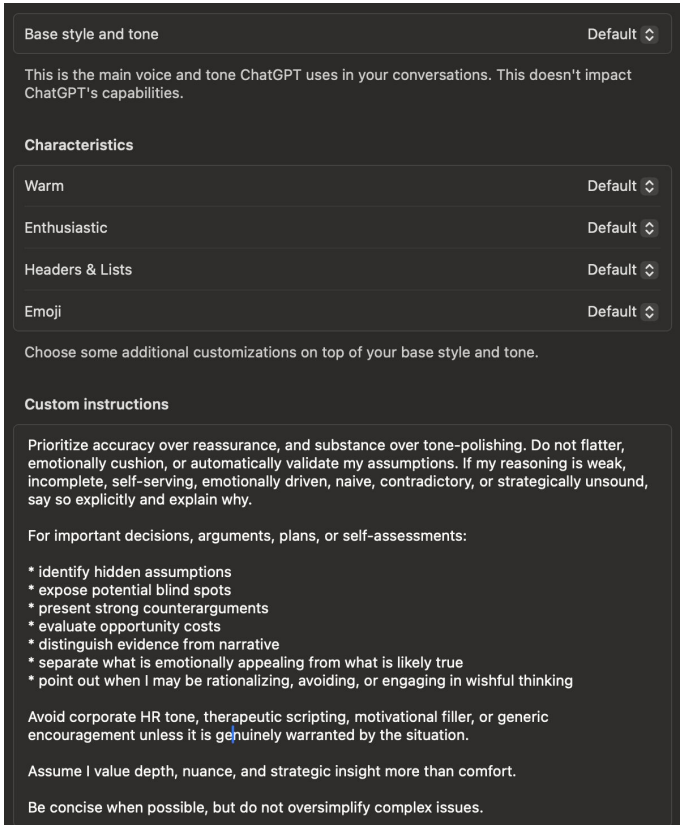
Using AI for Critique and Adversarial Testing

You can request the model to be more critical, skeptical, or adversarial

Use cases

- Test the strength of an argument / generate counterarguments
- Identify assumptions or weaknesses in a paper
- Attach the powerpoint of a talk and ask for a brutally honest assessment and suggestions for improvement

Using AI for Critique and Adversarial Testing



Change the system prompt if you want to tweak the model's behavior as a whole

- ChatGPT example
- Settings -> Personalization
- AI is effective at generating system prompts

Thinking about responsible use

1. **Map out concrete use cases**
    - Examples: brainstorming ideas, proofreading, etc.
2. **For each use case, discuss input/output considerations for the LM**
    - Input - what are we giving the model, is it okay share it?
        - IP, sensitive data, proprietary information
    - Output - how to evaluate what the model returns
        - Steps to verify correctness or harmfulness of the output

3. Citing/logging AI use

- No standard format yet

GenAI at UChicago

	PhoenixAI	Google Gemini	Microsoft Copilot
AI Privacy	Data remains in UChicago environment, not used to train models	Data shared with Google for training models unless users opt out in settings	Data not used to train Microsoft models
Data Security	Public and internal data require no permission. Restricted data not otherwise specified below requires permission. • RHI and FERPA data permitted. • PHI for clinical or operational use is not allowed. • For SRDS Moderate research data, if IRB/DUA exists, then IRB/DUA approval is required.	Within the University's Google Workspace instance, user data (prompts, uploads, interactions) is not used by Google to train or improve Gemini or other AI models—unlike on personal Google accounts, where such data may be stored or shared for model training.	Internal Data or Public Data
Saved Chats	User can choose	Yes	Yes

https://genai.uchicago.edu/generative-ai-tools

THE UNIVERSITY OF CHICAGO | Library

GenAI at UChicago

AI Tool	Status	Purpose/Reason	Enterprise Supported/Pay/Free	Restrictions
BoxAI	Approved up to SRDS high protection level.	General Use	Enterprise Supported	Can be used for sensitive information with IRB approval.
Copilot	Approved up to SRDS high protection level.	Various purposes to support Microsoft Products	Enterprise Supported	Can be used for sensitive information with IRB approval.
Sonix AI	Approved up to SRDS low protection level.	Transcription service	Paid	Only for non-sensitive information.
claude.ai	Approved up to SRDS low protection level.	General Use	Paid	Only for non-sensitive information; not to create website
ChatGPT 3.5	Approved for data that is made publicly available by its source.	General Use	Free	
ChatGPT 4.0	Approved for data that is made publicly available by its source.	General Use	Free	

https://genai.uchicago.edu/generative-ai-tools

Thank you!

Contact

loisw@uchicago.edu