Using AI in Research and Writing

How AI can help and hurt you at every stage of the research lifecycle, with special emphasis on MA theses

By Lois Wong, AI Librarian

Workshop Background

DiRT: Thesis Edition (**Di**gital **R**esearch **T**oolbox)

- CDS workshop series to support MA students, focusing on thesis support
- ~3 workshops per quarter

This is the first ever workshop! Would love your feedback for future iterations :)

Agenda

1. Introductions / Icebreaker
2. AI in the Research Lifecycle
3. Overview of AI Tools
4. Best Practices while Using AI
5. GenAI Resources at UChicago

Introduction

- Lois Wong
- AI Librarian since Sep 2025
- BA Linguistics, UC Berkeley
- MS Computer Science, Johns Hopkins
- Finished my master's project Dec 2024

Gaita, the face of my RAG chatbot that creates personalized CS Learning pathways from open courseware

Ice Breaker

Go around the room and share your

- Name
- Where you're from
- Current/past degrees
- Areas of interest or goals for your thesis (or capstone)

AI in the Research Lifecycle

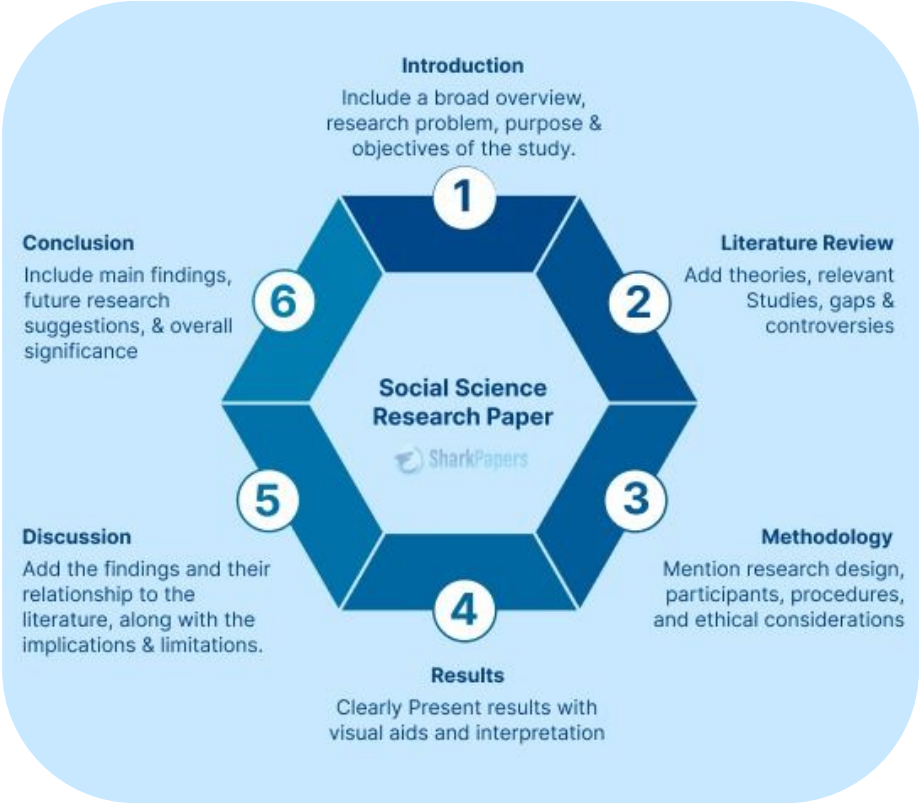
AI in the Research Lifecycle

How AI can help you at each stage and risks to be aware of

1. Snapshot of research lifecycle
2. Step through each stage and identify where AI can help and potential risks

The Research Lifecycle

1. **Observation** - Find topic of interest
2. **Lit Review** - read about your area of interest and find a niche
3. **Find available data/methods**
4. **RQ & Methods Refinement**
5. **Do the research (not covered)**
6. **Writing and Editing**



Discovering topics of interest

Where you can use AI

- Brainstorming - AI Assistants/chatbots as a sounding board

But at what (potential) cost?

- Suggestions may be shallow and reinforce common trends/biases
- IP risk: will cover later

Literature Review

Where you can use AI

- Deep Research to find sources
- Summarize large amounts of text
- RAG (e.g. through Phoenix AI)
    - Connect data source(s) and ask which page of this paper or which source discusses x topic
    - Great for papers you've read before

But at what (potential) cost?

- Your perspective is influenced by the unique path you chart through information/research
- Hallucinated sources / wasted time

Finding Available Data and Methods

Where you can use AI

- **Data Augmentation** - "artificially generating new data from existing data" to increase data diversity and quantity (for cases where you have less data than you need)
- **Data Imputation** - filling in missing values in datasets

But at what (potential) cost?

- Augmented data could obscure original patterns or introduce noise or false patterns

https://aws.amazon.com/what-is/data-augmentation/ ; https://milvus.io/ai-quick-reference/what-are-the-limitations-of-data-augmentation

RQ Refinement

Where AI can Help

- Recommending alternative phrasing, related questions

But at what (potential) cost?

- LMs reinforce the perspectives they were trained on, may reinforce biases/dominant perspectives
- Overuse can make your work seem generic

Examples of AI in Methods Sections

Topic Modeling: Find themes/topics in large amounts of text data (like Twitter), Uthirapathy et al., "Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA"

Word embeddings: Bolukbasi et al. finds gender bias in word embeddings in "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings"

Classification: Buolamwini et al. shows that facial gender classification has the highest error rate among darker-skinned women and lowest among lighter-skinned men in "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification"

Simulation

- Slater et al.'s "A Virtual Reprise of the Stanley Milgram Obedience Experiments" uses simulation to do what would otherwise be unethical
- An experiment to see how people would react shocking to virtual humans

Writing and editing

Where AI Can Help

- Proofreading and rewording
- LaTeX support

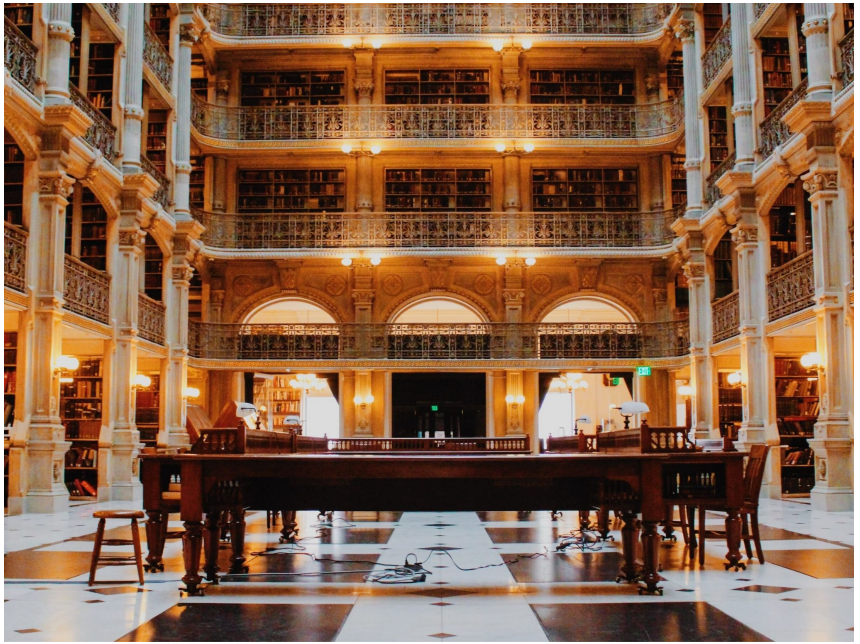
But at what (potential) cost?

- Over reliance on AI can compromise idea generation and connection formulation between concepts
- ‘Loss of your voice’ if overused
- Potential for academic violation - check with your preceptor

arXiv stopped accepting review and position CS papers due to the influx of AI generated content lacking in sophisticated analysis and synthesis

What each AI tool is good at

- **Gemini:** deep research, lit review
- **Perplexity:** good for looking things up (*given my interests, recommend some sources*). Deep research mode gives more peer reviewed papers.
- **Claude:** coding
- **ChatGPT:** writing, tends to be sycophantic
- **Notebook LM:** organizing notes (generate summaries, flash cards, podcast)
- **Phoenix AI:** connect data sources for RAG (e.g. papers you're citing in a Box, GDrive, or OneDrive folder)



Best Practices for Using AI

What Happens When You Use AI Tools?

- "When the service is free, you are the product" (source unknown)
- Your prompts and uploaded data are (usually) stored, analyzed, and used to further train AI models
- This exchange is what allows companies to offer powerful technology at minimal to no cost
- Be careful with sensitive, proprietary, and personal data: never assume privacy and confidentiality

Whenever you use AI tools, it's important to realize that anything you share, whether it's original ideas, documents, or creative work, might be incorporated into future versions and AI outputs.

What Happens When You Use AI Tools?

- As a response to those concerns among others, UChicago has taken some steps to protect our community's data while still providing access to helpful AI resources
- How many of you have heard of Phoenix AI?
- Phoenix AI is a Walled Garden of the same models that power ChatGPT

phoenixai.uchicago.edu

What is a Walled Garden?

- Secure and controlled environment where AI tools can be used safely
- Any prompts/data you input does not leave UChicago
- Further trained on a curated dataset to promote safety and accuracy
- Allows us to reap the benefits of AI without exposing ourselves or our data to unnecessary risks

AI Tools at UChicago

UChicago negotiated contracts and license agreements with these vendors (similar to enterprise plan). If you access these services through your UChicago account, you will get the additional data and privacy protections that don't come with the free plan.

- **Phoenix AI**
- **Microsoft Copilot**
- **GitHub Copilot Business**: Coding Assistant
- **NotebookLM:** Study tool, access through UChicago workspace

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

Best Practices for External AI Tools

- Review platform policies when uncertain especially before sharing valuable or sensitive data
- Mask sensitive data: use placeholders instead of actual data (e.g., “[Client Name]”)
- Treat every prompt like it could be seen by someone else

Recent oversight: ChatGPT conversations were indexed by Google and appeared in search results last year (summer 2025)

Best Practices for Responsible AI Use

How do we effectively and responsibly use AI?

- **Verify everything**, no matter how confident the LM seems
- **Cite/log AI usage**: always be prepared to answer questions on how you use(d) AI in your work
- **Keep original drafts** if you use AI to proofread your papers
- Always check with your professor/boss re. appropriate AI usage
- Rule of thumb: ChatGPT uses **10x more energy** as Google Search

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

2 more workshops this quarter

DiRT: Thesis Edition (Winter 2026)

1. **Introduction to Data Analysis and Visualization with Python** - *2/12 at 11am*
2. **Working with Secondary Data: Finding, Cleaning, and Reusing Existing Datasets** - *2/26 at 11am*

Thank You!

Questions?