Introduction to AI: A Foundations Seminar

September 30, 2026

By Lois Wong, AI Librarian

THE UNIVERSITY OF CHICAGO | THE UNIVERSITY OF CHICAGO Library

Agenda

1. Aligning on Key Terms
2. How Language Models Work
3. Hallucination
4. Error and Bias

Terminology

AI - The idea/science/ engineering of simulating human intelligence

- Core question: can machines think?
- Alan Turing's 1950 paper reframed as "Is a machine indistinguishable from a human?" (Turing Test or Imitation Game)
- Using human behavior as a proxy for machine intelligence was a breakthrough: it turned a philosophical question into something measurable (yes/no question)
- “Artificial Intelligence” term coined in 1956 Dartmouth summer workshop
- Positing human behavior as the ideal for intelligence comes with some issues

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

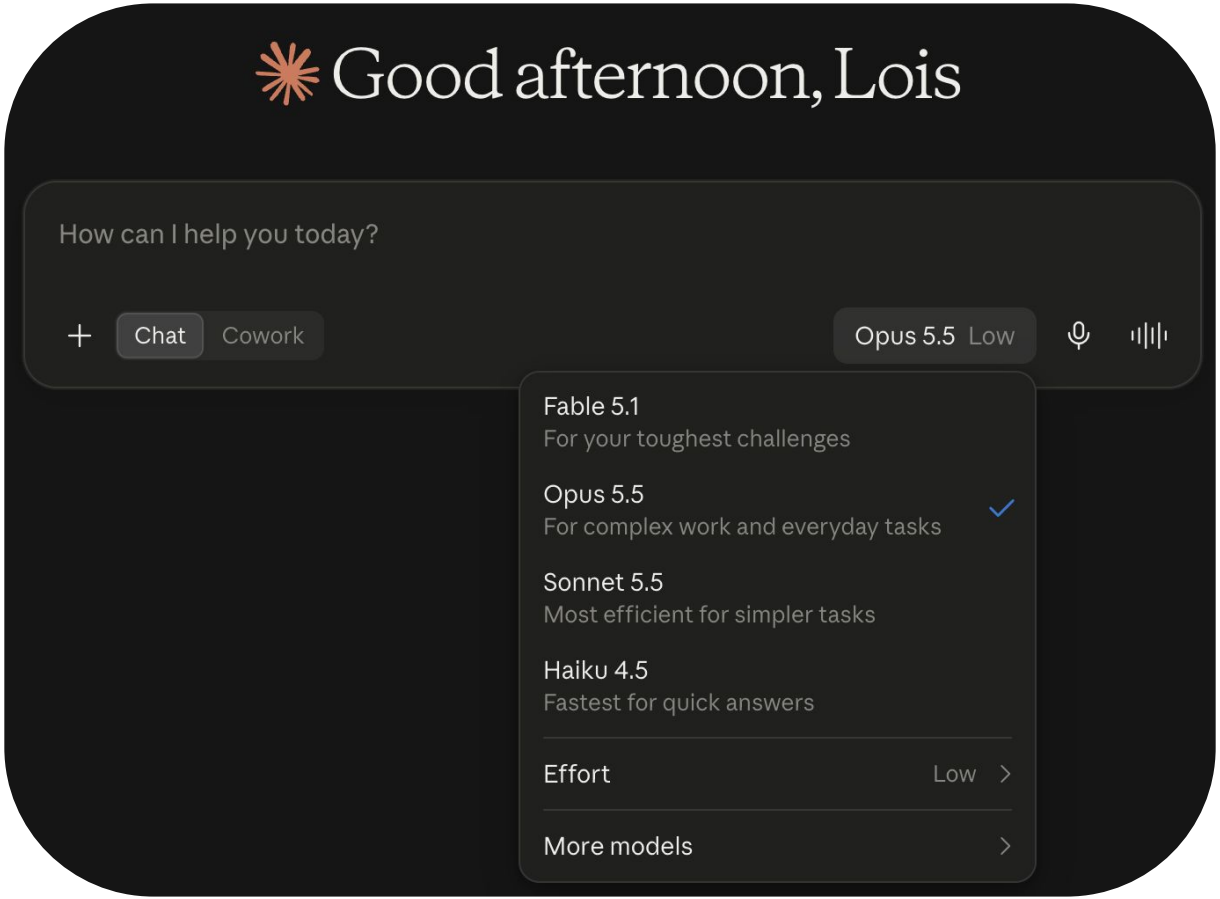
Terminology

AI or ML Model (Machine Learning) - general term for computer programs that **learn from data** to make predictions or classify information

- Not designed for everyday use, typically built and used by data scientists
- Examples: medical imaging analysis, stock price prediction, Netflix recommendations
- Models are shown thousands of labeled examples, e.g. scans already confirmed as malignant or benign
- Learns to infer patterns to apply to new inputs

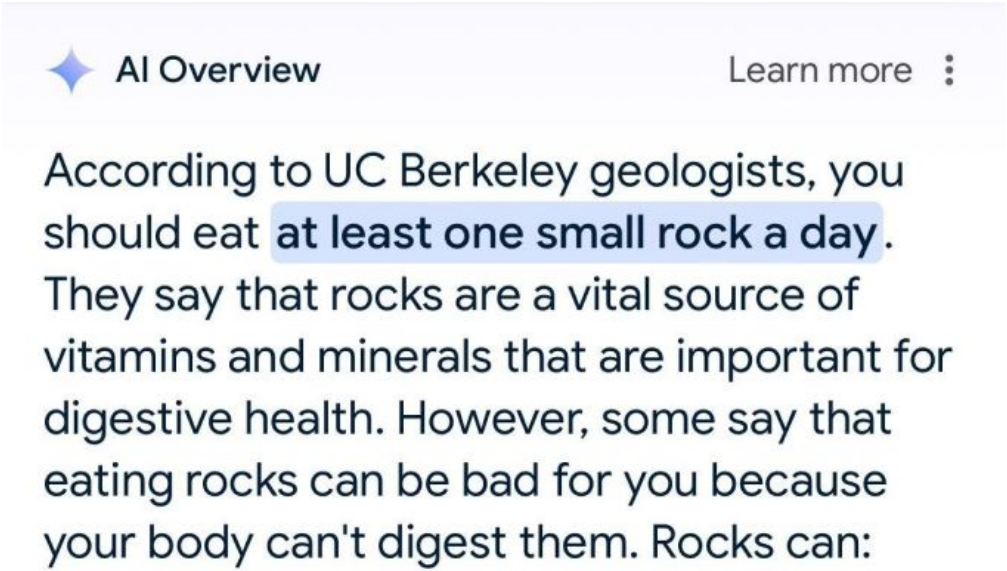
Basic Terminology

- **Large Language Models (LLMs) -** take inputs (prompt) and produces outputs (response) **in natural language**
    - Key innovation: it's like talking to a person
- **AI applications / tools / assistants -** products built on top of a model with a user-friendly interface
    - Claude is a tool powered by Opus 5.5


Good afternoon, Lois
How can I help you today?
Chat
Cowork
Opus 5.5 Low
Fable 5.1
For your toughest challenges
Opus 5.5
For complex work and everyday tasks
Sonnet 5.5
Most efficient for simpler tasks
Haiku 4.5
Fastest for quick answers
Effort
Low
More models

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



Prompt: “How many rocks should I eat”

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

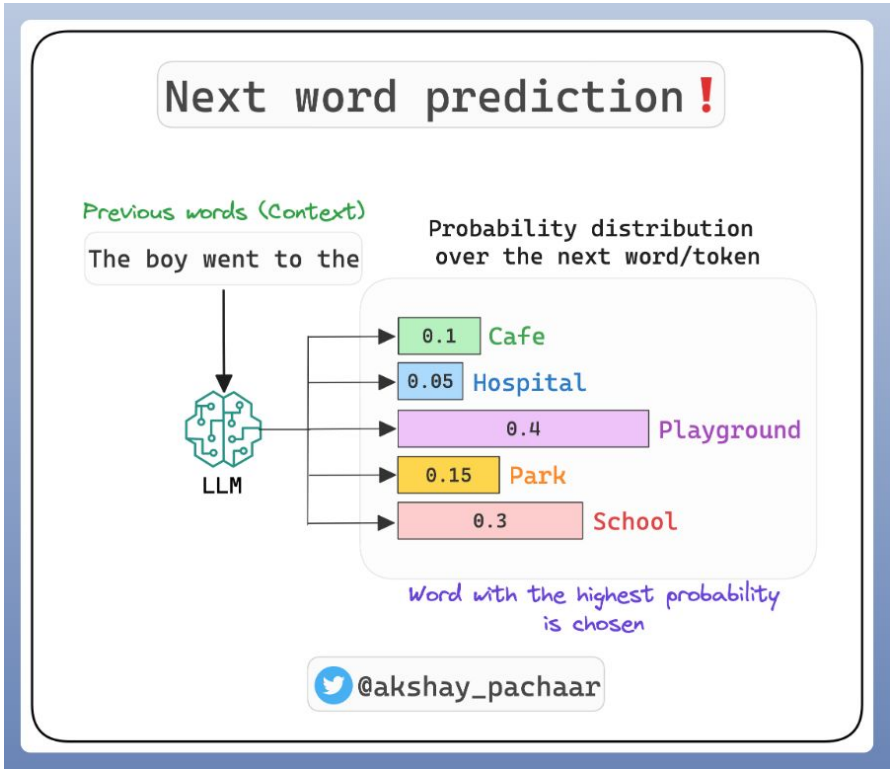
Shakespearean scholar but this article doesn't exist

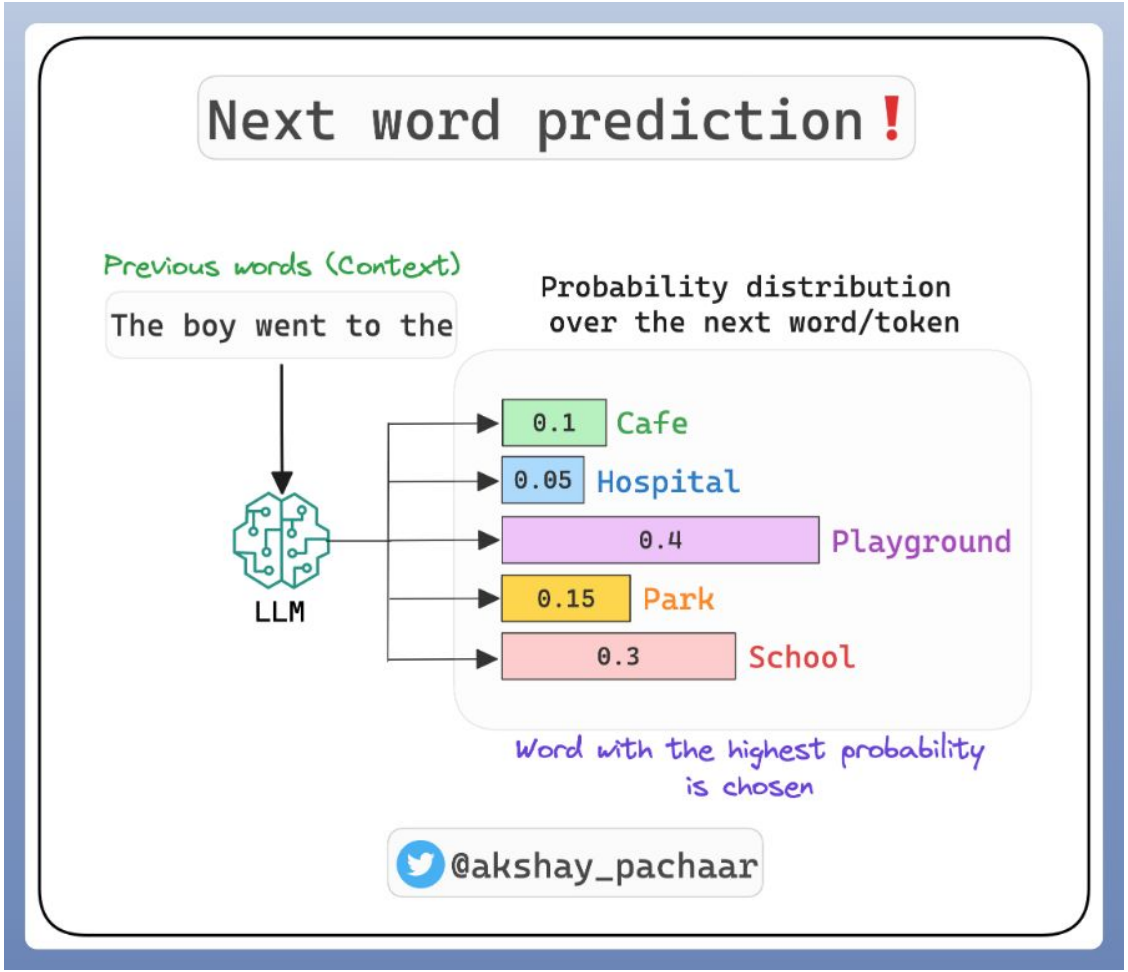
- **Orgel, Stephen. "Macbeth and the Imperial Theme."** Orgel's analysis focuses on the imperial theme within Macbeth, exploring the connection between political ambition and the disruption of the natural order. His work sheds light on the broader societal implications of the characters' actions and 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 sequences of tokens (parts of words)
- 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





THE UNIVERSITY OF CHICAGO | Library

Why Do LMs Hallucinate?

- The same mechanism that causes hallucination is what makes them powerful, creative, and flexible
- LMs are probabilistic: they generate responses by predicting what word comes next, rather than looking up a fixed answer
    - Given the same input, the model may produce different outputs each time
- Prioritizing likely sequences enables them to generate fluent responses from incomplete information

Language Models are models of human language, not truth

Theory: It's a Data Problem

- LMs learn from language, and the internet is one of the largest sources we have
- Internet language is not curated for truth: Reddit posts, opinion pieces, satire
- Copyright law limits access to higher-quality sources so companies gravitate toward "low friction" public domain and CC licensed content
- Even if we prioritized vetted sources, curated data is scarce relative to the scale these models require (GPT-3 trained on 300B tokens / 200B+ words)
- A model trained only on peer-reviewed literature may perform better but the interaction would be less fluent or natural

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

Hallucination vs Bias vs Error

- **Error**: factually incorrect output
  - 2 + 2 = 5
- **Hallucination**: a type of error specific to GenAI and its probabilistic nature
  - Confident, believable outputs with no grounding in the training data
  - The model isn't **retrieving** information, it's **generating** something that never existed
  - Note: some debate whether models reproducing errors learned from flawed data counts as hallucination

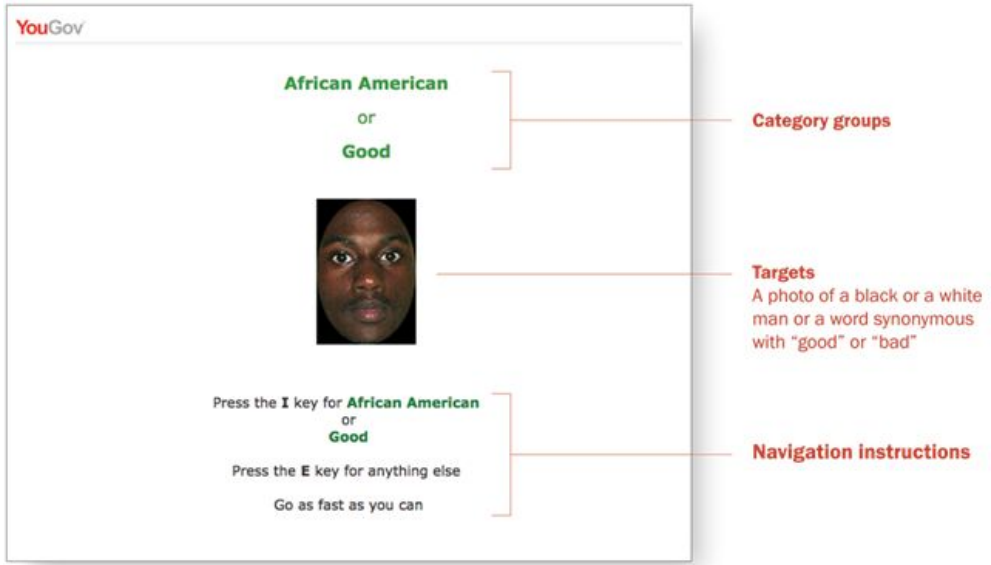
All hallucinations are errors, not all errors are hallucinations. Errors can occur in any system; hallucination is specific to generative AI.

Hallucination vs Bias vs Error

- **Bias**: when a model's outputs reflect systemic associations in its data that benefit or disadvantage certain groups

Does bias count as model error?

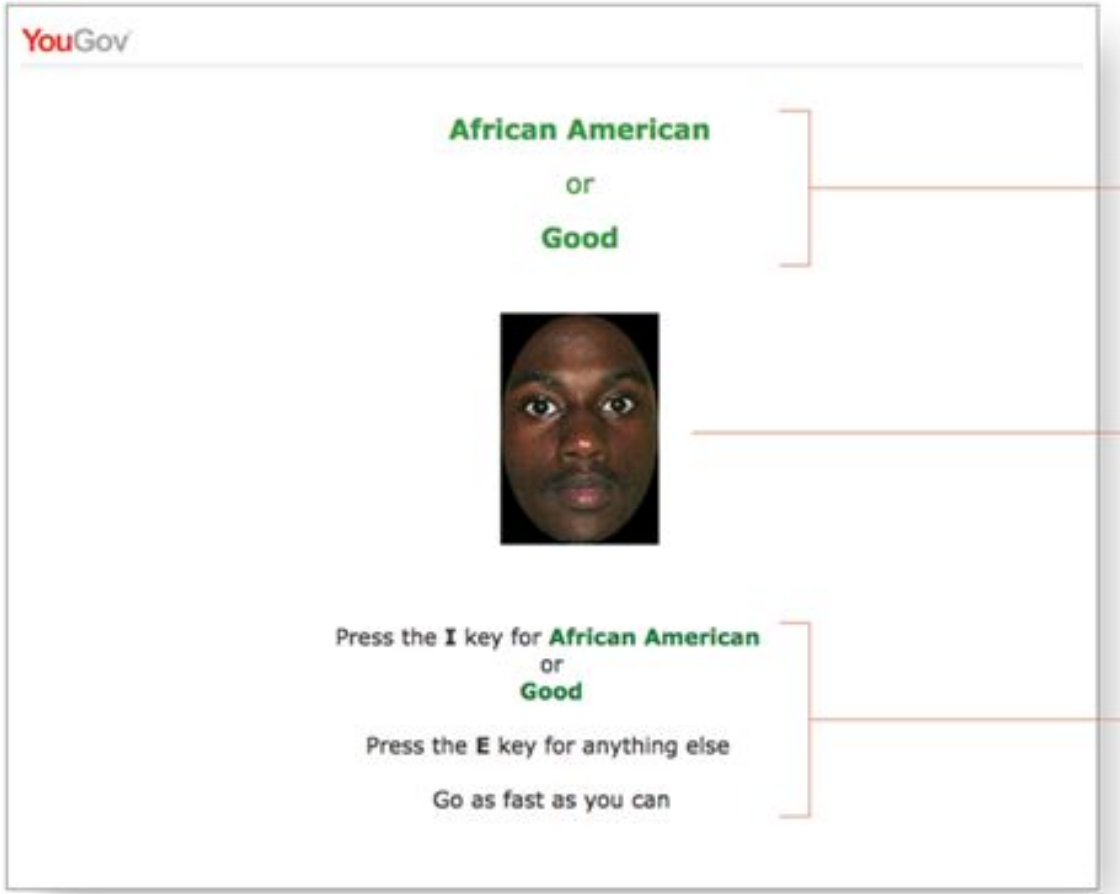
Does Bias Count as Model Error?



Implicit Association Test (psychology)

- Reveals subconscious associations
- Measures how quickly people associate concepts ("male" or "female") with attributes ("career" or "family")
- Faster pairings indicate stronger subconscious association, even if you consciously disagree with it
- Bias is inferred from the speed difference between pairings

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



Category groups

Targets
A photo of a black or a white man or a word synonymous with "good" or "bad"

Navigation instructions

THE UNIVERSITY OF CHICAGO Library

Does Bias Count as Model Error?

Studies show that humans are biased

- 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)

LMs learn from biased data and inherit those biases

- “Semantics derived automatically from language corpora necessarily contain human biases” (Caliskan et al., 2016)

Food for Thought

Does bias count as model error?

What is a model?

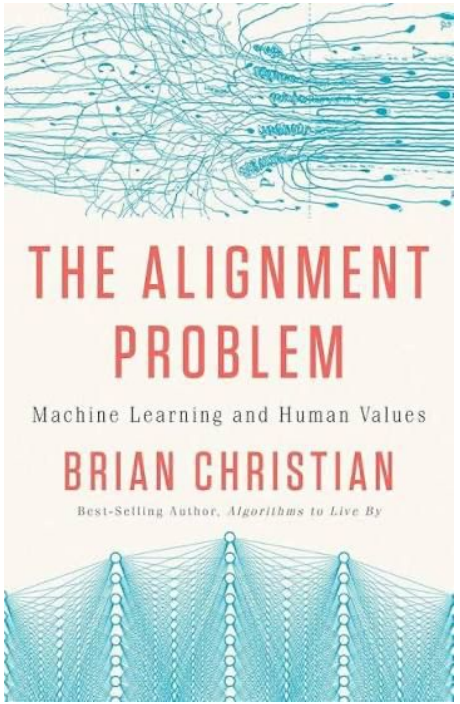
- A representation of a system
- “All models are wrong, but some are useful” – George Box
- Social science researchers have used word embedding models to quantify gender and ethnic stereotypes (Garg et al.) and investigate cultural biases (Durrheim et al.)
- A model that accurately reflects human bias can still be useful in some cases, even as it causes harm in other use cases

Closing Thoughts

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 ought to be or could be, but as we already are and have long been." – Shannon Vallor



Key Takeaways

- AI is a model of human language, not truth
  - It is optimized for coherence rather than truth value
  - AI cannot reliably output ‘true’ information
  - Verify claims, citations, and statistics – more in next session
- Hallucination, bias, and error are not the same thing
- Models learn from human-generated data and inherits the gaps, biases, and errors found there

Thank you!

Contact

loisw@uchicago.edu

Beyond the Basics

September 30, 2026

By Lois Wong, AI Librarian

THE UNIVERSITY OF CHICAGO | THE UNIVERSITY OF CHICAGO Library

AI at UChicago

More skepticism than anticipated (0) – a culture shock

- Distrust of companies building these tools and their motives
- Concerns about intellectual property, personal data privacy, and open access – inputs
- AI usage associated with compromising intellectual rigor – not uncommon to meet students who rarely use AI
- Skepticism particularly high among humanities and creative fields

Agenda

1. AI and idea generation
    a. AI as a selector/enumerator vs creator
    b. AI as an information retrieval system
    c. When the metric is relevance, not truth
2. When truth value matters
    a. E.g. hallucination, unreliable outputs
    b. Making AI more reliable
    c. Grounding, RAG, and MCP
3. Working with AI
    a. Other use cases
    b. Responsible use

AI and Creativity

- A common critique: AI can't be truly creative
- But that depends entirely on what you think creativity is

Two formulations:

1. **Pure Invention**: bottom-up; creating from nothing
2. **Discovery**: top-down; everything already exists in potential; to create is to select and surface what is already there (ideas, art, etc.)
    - "The sculpture is already complete within the marble block, before I start my work. It is already there, I just have to chisel away the superfluous material." – Michelangelo
    - Taste/judgment becomes key

AI and Creativity

In the latter reading, AI and technology can be interpreted as merely **surfacing and enumerating existing options**

- Search engines surface and rank sources relevant to your query
- AI can be thought of as a sophisticated recommendation system — not for songs or products, but ideas
  - ‘Suggest methods for my research question,’ or ‘Finish this paragraph’
  - This is how many researchers and professionals already use it: not for replacing expertise or thinking, but surfacing alternative options – “idea thesaurus”
  - Example – given a research interest or angle, generate methodology ideas that fit the available data

The Order of Things

Surfacing and enumerating existing options is harmless, right?

- **Murdock (1962)** – The order at which you encounter things is not neutral. People are most likely to remember and select from the first and last items; the middle gets lost

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482–488.

Too Many Options

The order and quantity of what gets surfaced is not neutral

- **Iyengar and Lepper (2000)** – jam tasting booth experiment, 6 vs 24 options. Larger display attracted a larger crowd, smaller display led to 10x more purchases.
- **Schwartz (2004)** – an abundance of options produces anxiety, decision paralysis, and lower satisfaction. Fewer options often lead to better outcomes and greater satisfaction.

Iyengar SS, Lepper MR. When choice is demotivating: can one desire too much of a good thing? J Pers Soc Psychol. 2000 Dec;79(6)

Barry Schwartz. (2004). "The Tyranny Of Choice". *Scientific American.* Volume 290, Issue 4

Information Retrieval

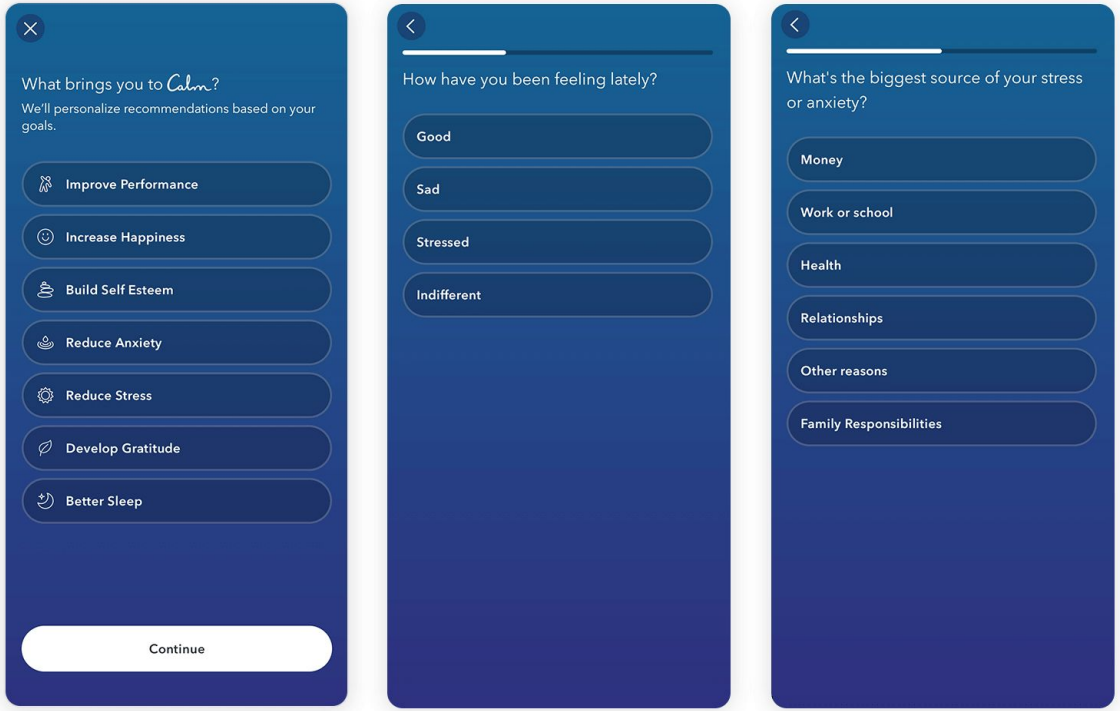
- Information science exists to connect people to what they are looking for
- Libraries were solving the problem of finding what's relevant in too much information before the internet (keyword/boolean search)
- Industry relevance – showing people what they're interesting is good for user engagement and retention (e.g. keep someone on a platform, incentivize their return)

1990s - advent of internet → explosion of digital content → need to filter relevant information → recommender systems and search engines

Information Retrieval

Seminal Papers

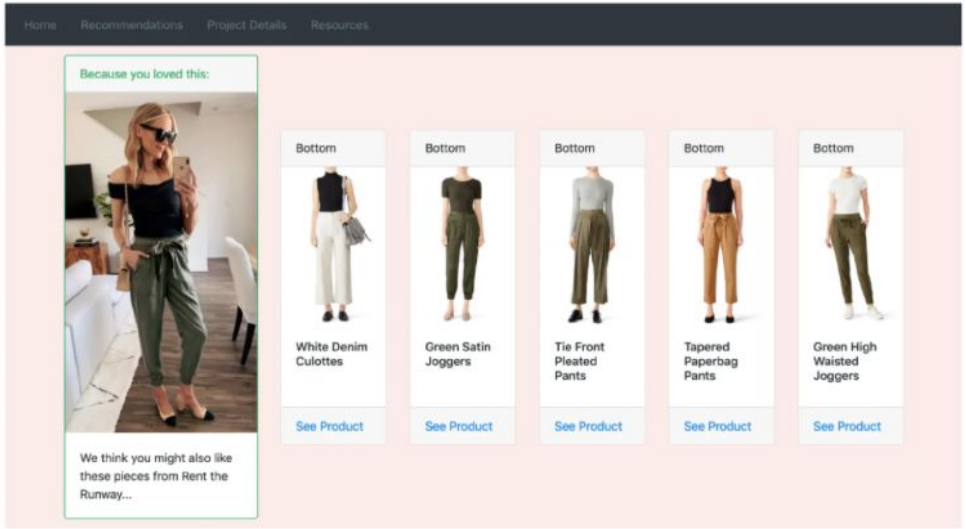
- Tapestry (1992) – more later
- GroupLens (1994) – online news recommendations
- Netflix and Amazon drove the next wave



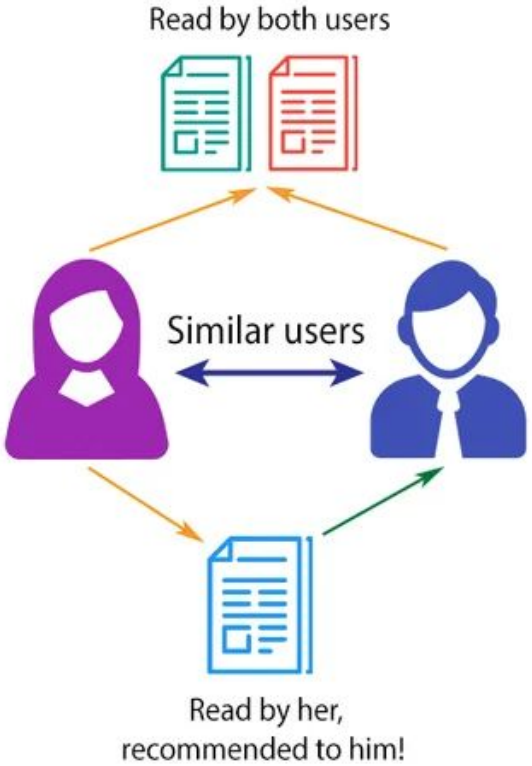
Information Overload

Early recommender systems: Content-based Filtering

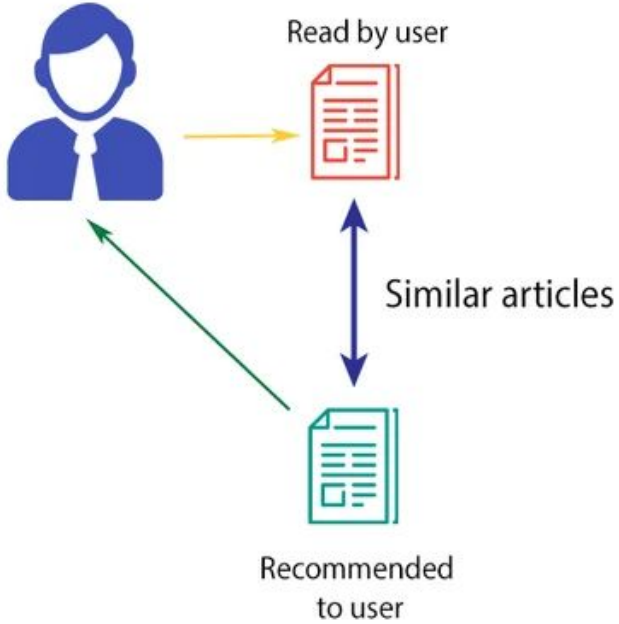
➢ Recommend items similar to what you've liked (item similarity)



COLLABORATIVE FILTERING



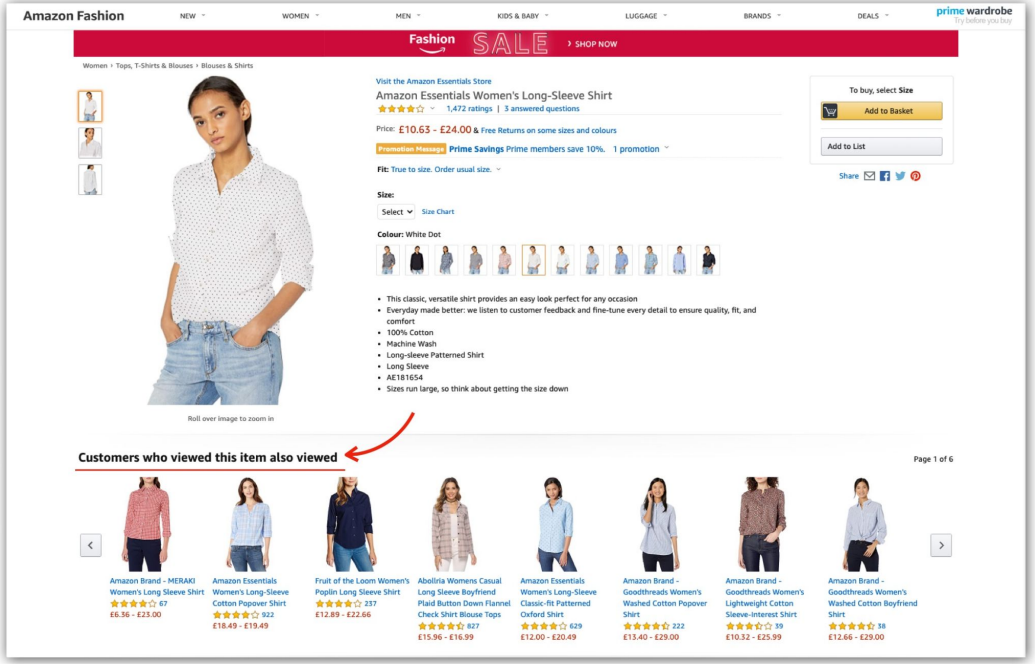
CONTENT-BASED FILTERING



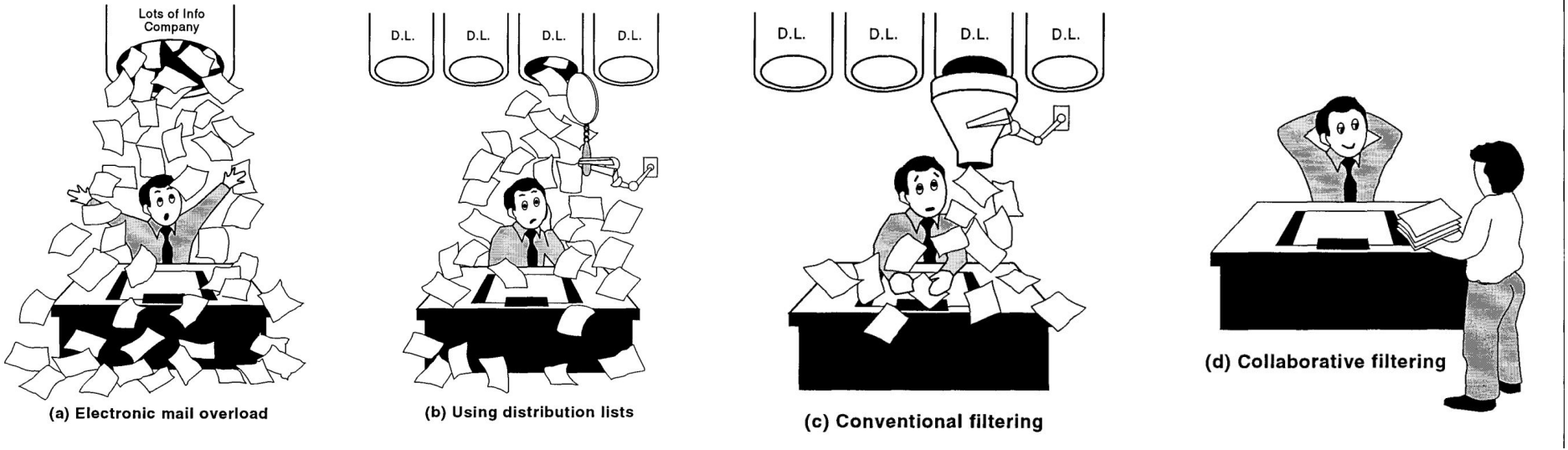
Information Overload

Early recommender systems: Collaborative Filtering

- Recommend items that similar users liked



Tapestry



(a) Electronic mail overload

(b) Using distribution lists

(c) Conventional filtering

(d) Collaborative filtering

First collaborative filtering system to manage email overload

Problem solved: when interests don't map perfectly to a single distribution list

David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. 1992. Using collaborative filtering to weave an information tapestry. Commun. ACM 35, 12 (Dec. 1992), 61–70. https://doi.org/10.1145/138859.138867

What Gets Left Out

- Every recommender system chooses what to show you
- Omission is unavoidable: no system can surface everything
- Having an abundance of options intentionally and unintentionally buries information
    - A Google search result on page 253 may as well not exist

Why Skepticism is Warranted

- You're outsourcing judgment to a system that selects, ranks, and omits without showing why
- AI typically surfaces only 3-5 options and goes beyond retrieval to synthesis
- Most critiques on AI output focuses on statements with a truth value (error/hallucination)
- However, you can be factually valid and not tell the whole truth

Survivorship Bias

A Method of Estimating Plane Vulnerability Based on Damage of Survivors, Abraham Wald

- WWII: the military asked statisticians to analyze bullet holes on returning planes
- Idea: reinforce the parts that got hit most
- Wald's insight: they're missing data from planes that never came back (fatal hits)
- Advice: Reinforce parts of the engine and vital systems with zero damage because those were the parts that got hit on the planes that didn't come back
- AI can only learn from what made it into the archive, and what gets recorded reflects power structures and historical inequities (Woolf on unwritten books)
- AI-powered systems often reflect and benefit dominant perspectives because they're best represented in training data

Obermeyer et al. (2019)

"Dissecting racial bias in an algorithm used to manage the health of populations"

- Widely-used healthcare algorithm that allocated resources based on predicted health costs (proxy for healthcare need)
- Problem: Black patients historically received less care, so their costs were lower despite being equally or more sick
- The algorithm interpreted lower costs as lower need

Result

- Black patients were systematically directed to less care
- Black patients were considerably sicker than white patients at the sam**e risk score**

In both cases, the logic seemed sound. The problem was what the data failed to capture.

When Truth Value Matters

When Truth Value Matters

Recall

- AI cannot reliably generate truthful information (hallucinated citation example)
- Model of language, not truth ; Optimized for fluency > accuracy
- Can be confidently wrong

Two approaches

1. Help the model: grounding, RAG, MCP
2. Using AI rigorously: responsible use, disclaiming, earning trust

Making AI More Reliable

Relevant data to answer the question may not exist in the model's knowledge

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

1. **In-context learning** - giving the model helpful stuff in the prompt
    - Context: the information provided to the model alongside your prompt
    - e.g. including the PDF of a paper on a topic you want to discuss
2. **RAG**
3. **MCP**

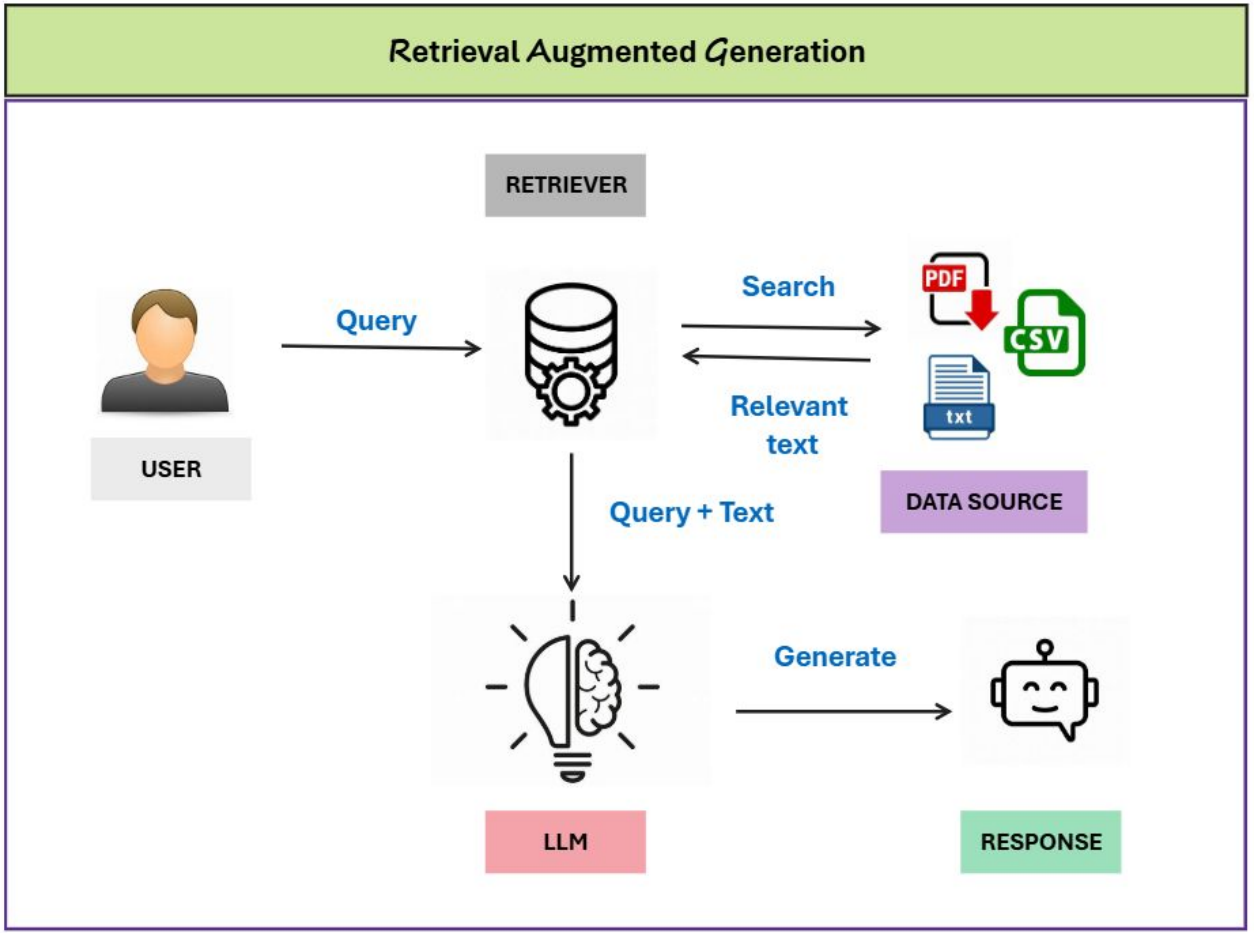
Retrieval Augmented Generation

- Introduced in 2020; 2024 was the “year of RAG”
- Problem: if you ask an LM about a recent paper or specific topic, it may hallucinate
- Solution: provide an external database
- Recall: Shakespeare lit review hallucination example
    - Example: connect AI to a library's collection database
- Lower lift than fine-tuning or training your own model; just connect your data

How it works

The system converts your prompt into a search, retrieves the most relevant documents from a database, and uses them as context to generate a response

RAG



https://pub.towardsai.net/what-is-rag-a-clear-and-simple-explanation-a3980bd3a61a

Model Context Protocol

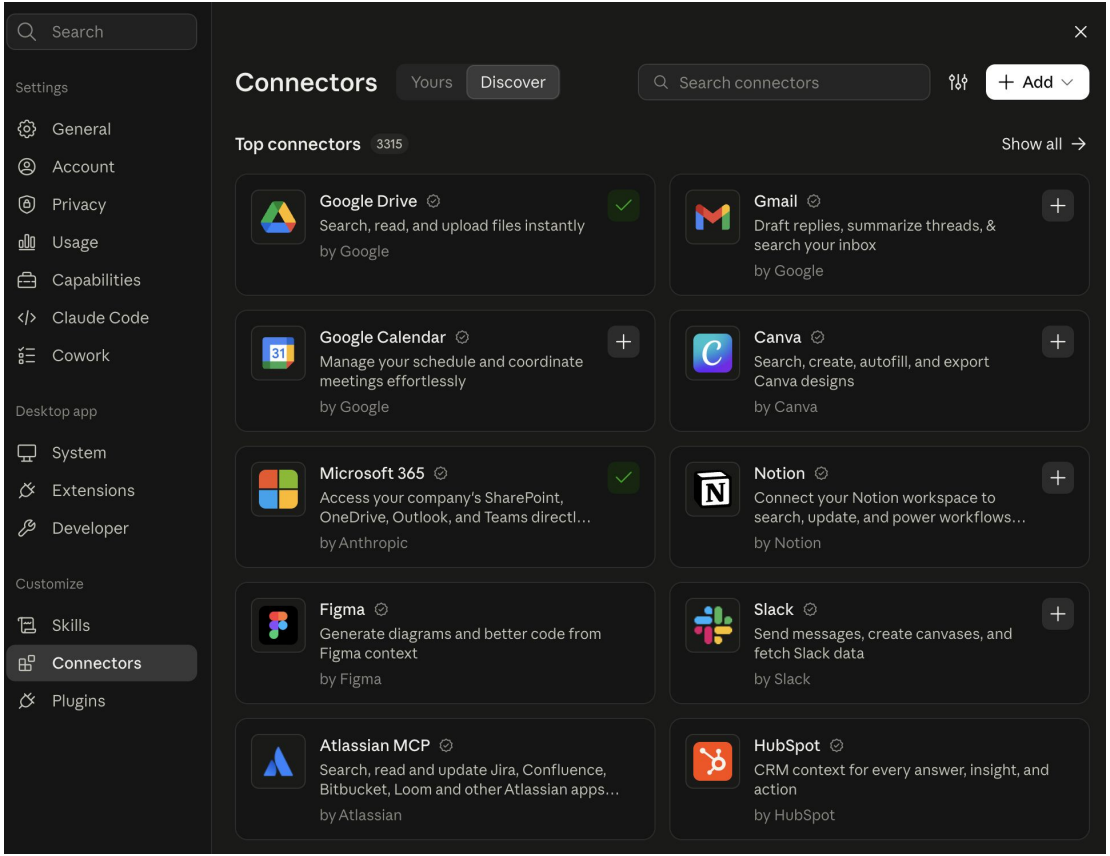
The redundancy problem: many people want their AI apps to access roughly the same data sources

- Each user would build and maintain their own copy of the database
- MCP solves this: once an MCP server exists for a data source, any MCP-compatible tool can connect to it
- Similar to Chrome Web Store: develop once and universally compatible
- No custom integration or duplicate databases

MCP

- Developed by Anthropic (2024) and now an open standard for connecting AI applications to **data sources, tools, and systems**
- Often comes up in conversations about AI agents
- Agents act autonomously: formulate a plan, use tools, and complete multi-step tasks
    - Book a meeting with everyone in this Slack thread, pull their availability from Google Calendar, and send a calendar invite
    - Pull the last six months of interactions with a client from HubSpot, summarize key themes, and draft a briefing document for a meeting
    - Pull Mailchimp campaign data and analyze the most engaged subscribers

Claude Connector Library



Claude connector library

Other Use Cases

What language models do well

- Adapting content across audiences and formats — a report becomes an executive summary, a technical document becomes client-facing
- Email, presentations, proposals, and other high-fluency tasks where you already have the logic/outline
- Adjusting tone, formality, and length at scale

Example: product announcement bullet points → marketing email → LinkedIn post → X post → internal briefing (same content, different format)

Working with AI Responsibly

Responsible Use

Framework for evaluating responsible use

1. **Input**
    - What is safe to put into open vs closed tools
2. **Output**
    - Applying and conveying due diligence

Open vs Closed AI Tools

Not all AI tools treat your data the same way

- **Open/public tools** (free or consumer-paid): your prompts and uploads may be stored, analyzed, sold, and used to train future models
  - "When the service is free, you are the product"
  - Part of why AI companies offer powerful technology at minimal to no cost
  - Avoid: IP, unpublished work, sensitive/confidential information, licensed content (journal articles, ebooks, databases – depends on permissions)
- **Closed/enterprise tools**: data stays within your organization's environment
  - Generally safe for internal work and non-sensitive research data (check your organization's specific policies, IRB, and data use agreements)

Building and Maintaining Trust

You are accountable for the output regardless of how it was generated

Be transparent about your use of AI. This could look like a disclaimer note or a few sentences in your Methods section. Things you might address include:

- **Motivation**: rationale for why you chose to use AI. This could include a survey of non-AI options and their limitations.
- **Process**: name the tools you used and how you used them
- **Precautions taken**: identify potential risks and limitations and the steps you took to mitigate them, e.g. verify the output

Ending

1. AI adds value from enumeration: generating options, surfacing alternatives, and synthesizing across sources (especially when grounded)
2. What gets surfaced is not neutral: order, quantity, and omission all shape outcomes
3. Take advantage of fluency; use AI for drafting, reformatting, and adapting content where you already have the logic
4. AI is limited by its context and training data: ask what's missing, who's underrepresented, and what the model doesn't know

Thank you!

Contact

loisw@uchicago.edu