

Lecture 2

Using AI for Scientific Writing and Development

2026-04-21

Using AI for Scientific Writing and Development

- AI-augmented research workflow
- AI as cognitive infrastructure for research
- Prompting and interaction paradigm
- AI for literature review and knowledge extraction
- Grounding and RAG systems
- Knowledge management and research workflows
- Scientific writing with AI
- AI in programming and research systems

Material is available at: <https://www.itn.liu.se/~siaih22/6fitn80.html>

AI-augmented research workflow

Literature → Extraction → Grounding → Writing → Coding → Evaluation

- AI acts as a co-pilot across all stages, not just a single tool
- Iterative nature of the workflow

Literature review & discovery

- AI for: semantic search across large corpora, rapid identification of relevant papers
- Shift from keyword search → meaning-based retrieval

Knowledge extraction

- Automatically extract key claims, methods, results
- Summarize and compare multiple sources

Grounding (RAG and verification)

- Connect model outputs to verified documents, scientific sources
- Goal: reduce hallucinations, improve factual accuracy

Scientific writing

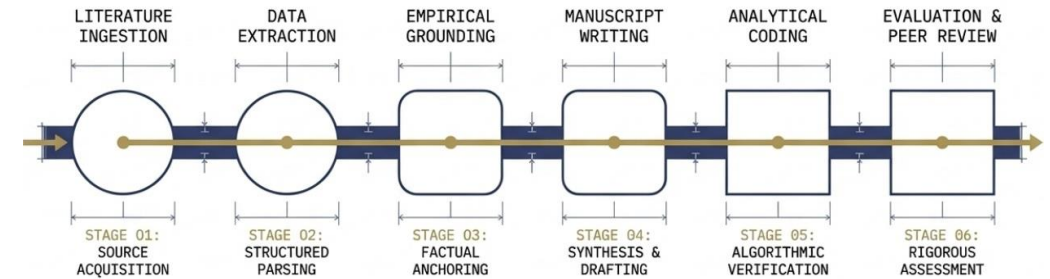
- AI assists with drafting sections, editing for clarity and style, structuring arguments
- Maintain human authorship and responsibility!

Coding & experimentation

- code generation, debugging, rapid prototyping
- data analysis pipelines

Evaluation & validation

- AI helps benchmark outputs, compare methods, automate evaluation
- Human verification remains essential



AI transforms research into an iterative workflow where literature, analysis, writing, coding, and evaluation are continuously supported and accelerated by intelligent systems

Role of AI as intellectual research assistant

Information management	<ul style="list-style-type: none">• Summarize and synthesize large texts• Extract structured insights from unstructured sources• Generate annotated literature reviews
Scientific writing support	<ul style="list-style-type: none">• Draft and revise sections of manuscripts• Suggest alternative phrasing and stylistic improvements• Check for clarity, coherence, and consistency
Hypothesis generation & idea exploration	<ul style="list-style-type: none">• Suggest novel angles or research questions• Compare theoretical frameworks across disciplines• Facilitate brainstorming through iterative prompts
Data analysis assistance	<ul style="list-style-type: none">• Assist with data preprocessing, cleaning, and transformation• Recommend relevant algorithms or statistical approaches• Generate interpretable summaries of experimental results
Retrieval-augmented grounding	<ul style="list-style-type: none">• Use RAG (Retrieval-Augmented Generation) to reduce hallucinations• Connect responses to credible sources, papers, or datasets
Coding & workflow automation	<ul style="list-style-type: none">• Automate routine coding or scripting tasks• Debug or review code snippets interactively
Critical thinking & verification	<ul style="list-style-type: none">• Provide alternative interpretations of data or text• Highlight potential inconsistencies or methodological issues• Encourage reflective evaluation rather than passive acceptance
Collaboration & knowledge sharing	<ul style="list-style-type: none">• Facilitate team-based research by summarizing discussions• Serve as a knowledge repository across multiple projects

Artificial Intelligence vs Augmented Intelligence in research

Artificial Intelligence

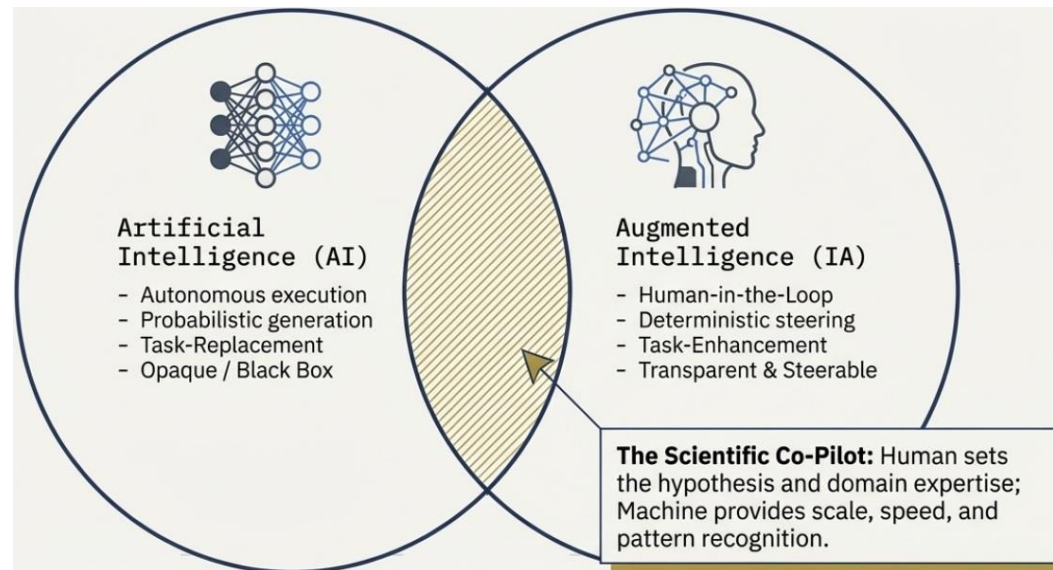
often implies systems performing tasks autonomously, potentially replacing human decision-making

- Focus on automation of tasks
- Systems aim to replicate human cognitive functions
- Often used in fully automated pipelines
- Risk of overreliance and reduced human oversight
- Typical framing in media and industry narratives

Augmented Intelligence

emphasizes human–AI collaboration, where AI enhances human cognitive capabilities rather than replacing them

- Focus on human–machine collaboration
- AI serves as a cognitive support system
- Enhances analysis, synthesis, and decision-making
- Keeps human judgment and accountability central
- Aligns better with scientific and research workflows



In research, the most effective use of AI is not full automation, but augmentation—combining machine efficiency with human judgment and expertise

From search engines to research co-pilots



Keyword-based retrieval of documents

User must:

- formulate queries
- filter results
- synthesize information manually

Limitations:

- information overload
- limited understanding of context and intent

Modern systems use:

- embeddings
- similarity search

Retrieve results based on meaning, not keywords

AI systems that:

- search
- summarize
- explain
- generate insights

Benefits for researchers

- reduced time spent on searching, filtering
- increased focus on analysis, interpretation, creativity

Risks and limitations

- potential hallucinations
- lack of source transparency (without grounding)
- over-reliance on generated summaries

Research tools are evolving from passive search engines into active co-pilots that assist with discovery, synthesis, and reasoning, fundamentally changing how research is conducted

Large Language Models as cognitive tools for scientists

LLMs as cognitive extensions

- memory
- reasoning

! But not to replace human cognition

Capabilities relevant to research

- text summary
- synthesis across multiple sources
- explanation of unfamiliar concepts
- hypothesis and ideas generation

From tools to thinking partners

Cognitive offloading

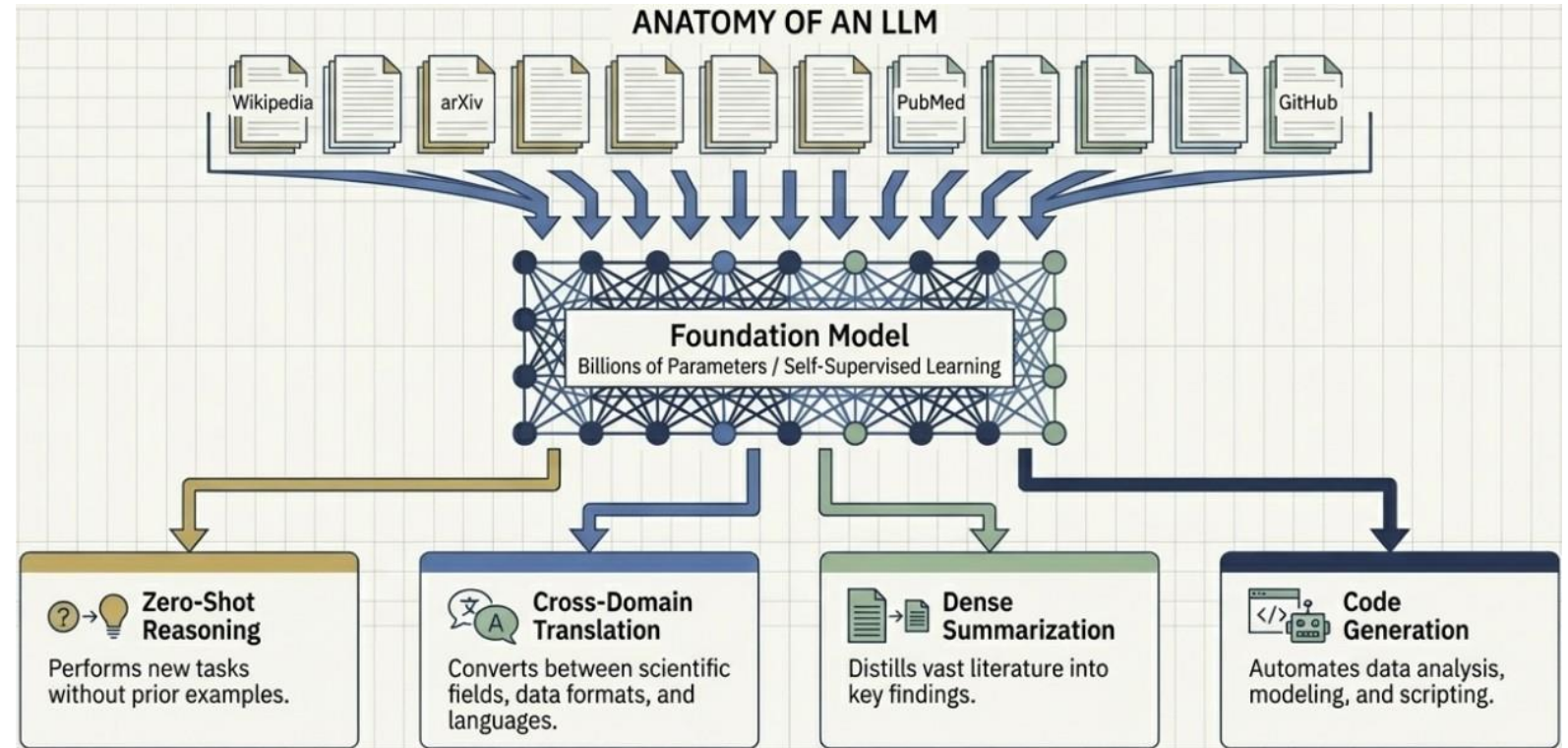
- reduce mental load for information retrieval, structuring ideas
- allow researchers to focus on critical thinking interpretation

Limitations as cognitive tools

- lack true understanding or grounding
- hallucinations, plausible but incorrect reasoning

Human–AI complementarity

- humans provide: domain expertise, judgment, validation
- AI provides scale, speed, pattern recognition



LLMs act as cognitive tools that amplify human capabilities, enabling scientists to process, generate, and reason about information more efficiently—while still requiring critical oversight

Sources: Brown et al, Language models are few-shot learners, arXiv:2005.14165v4

How LLMs process and generate scientific text

From text to tokens

- scientific text is first converted into tokens
- tokens may be words, subwords, punctuation symbols

Embeddings: numerical representation

- each token is mapped to a high-dimensional vector
- vectors capture semantic similarity, contextual relationships

Contextual processing with transformers

- tokens are processed through self-attention layers
- each token representation is updated using surrounding context
- crucial for: terminology disambiguation, method/result interpretation, citation context

Predictive generation

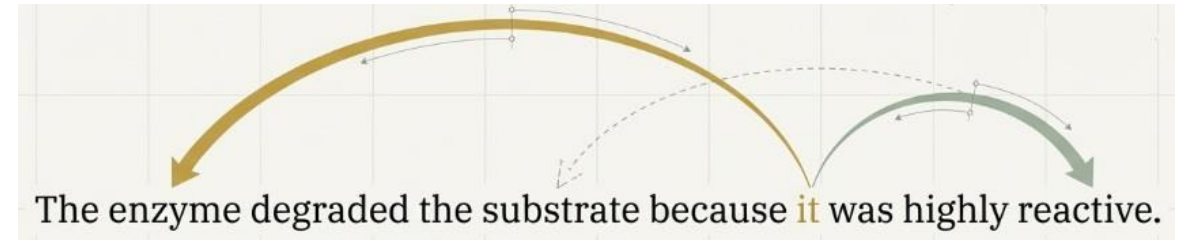
- LLMs generate text by predicting the next most probable token
- this continues iteratively until a complete response is produced

Scientific writing strengths

- coherent technical prose
- style adaptation to academic tone
- summarization and paraphrasing

Important limitation

- Models generate based on statistical likelihood, not factual verification
- Implication: fluency ≠ correctness



LLMs process scientific text by converting tokens into contextual representations and generating outputs through next-token prediction, producing fluent text without inherent fact-checking

Prompting as a new form of Human-Computer Interaction

New HCI model: graphical commands → dialogue-based interaction

Prompt as interface logic

- no longer just a question
- functions as instruction, task definition, contextual framing, output specification

Specific techniques

- chain-of-thought
- tree-of-thought
- RAG

Language as control mechanism

- “Compare the methods sections of these two papers and identify methodological differences.”

Interactive and iterative workflow

- Intent → Prompt → Response → Refinement → Re-prompt

Prompting as cognitive collaboration

- moves interaction beyond information retrieval
- supports brainstorming, critique, simulation of peer review, exploratory reasoning

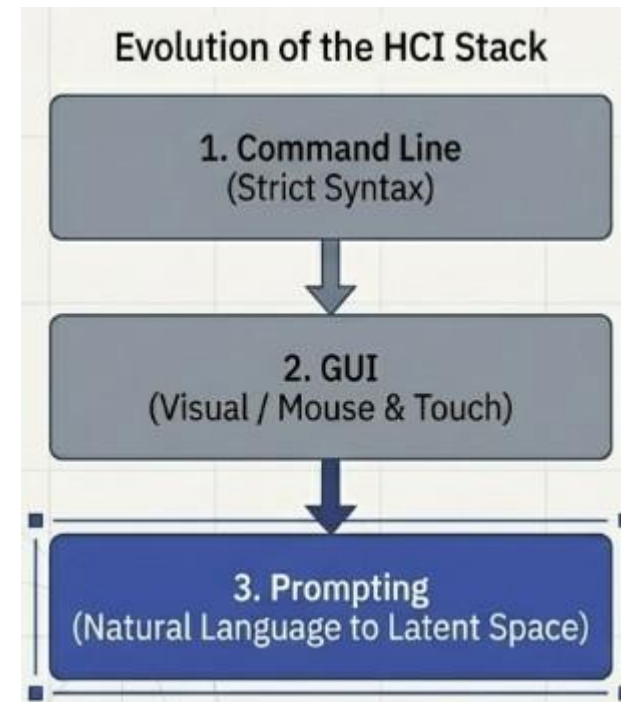
In-context learning as interaction

- model's ability to temporarily learn from prompts

HCI implications for researchers

- interaction quality strongly affects output quality
- prompt design becomes a research skill

⇒ How do we design human–AI interactions that are reproducible and trustworthy?



Prompting transforms human-computer interaction from static interfaces into an adaptive dialogue, where natural language becomes the primary mechanism for controlling and collaborating with AI systems

Prompt engineering for scientific tasks

What is prompt engineering?

- systematic design of prompts to obtain reliable, high-quality outputs from LLMs
- in research, prompts should be treated as part of the methodological protocol
- a prompt is an experimental instruction, not just a question

Core design principles

- Effective prompts should specify:
 - task → what to do
 - scope → what to include / exclude
 - format → expected structure
 - criteria → quality requirements

“Summarize the methods section in 5 bullet points, focusing on sample size, experimental design, and statistical tests.”

Prompt templates for scientific work

- Task + Context + Constraints + Output format

Few-shot prompting

Structured prompting for research

Verification prompts

- prompts should include self-check instructions

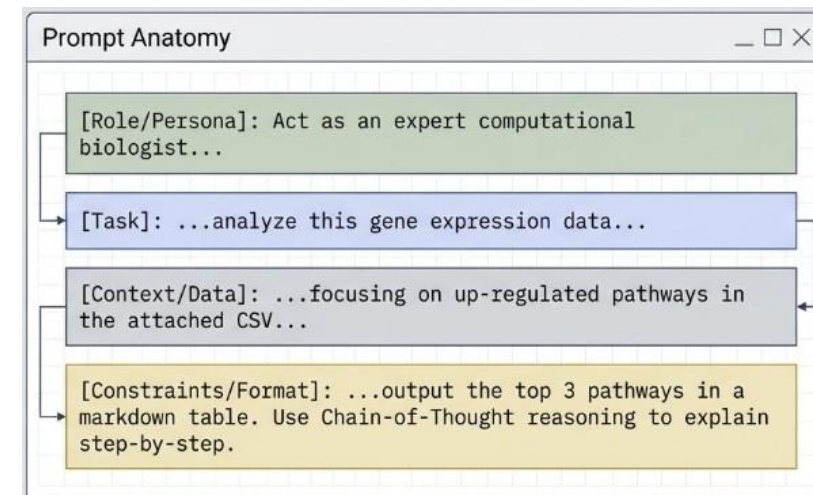
“State uncertainty where evidence is insufficient.”

“Cite the exact paragraph supporting each claim.”

Limitations

- outputs remain sensitive to wording
- prompt changes may alter conclusions
- requires human validation

Prompt engineering for scientific tasks is the methodological design of AI instructions that improves rigor, reproducibility, and reliability across research workflows



AI-assisted literature reviews at scale

Why scale matters

- researchers face information overload
- thousands of new papers are published across disciplines every week

Role of AI in literature review

- AI systems can support:
 - large-scale paper discovery
 - rapid screening of abstracts
 - clustering by topic and method
 - extraction of key findings
- this shifts effort from searching → synthesis

Semantic search and retrieval

- AI uses embedding-based semantic search
- finds papers by meaning and conceptual similarity, not only keywords

Automated summarization

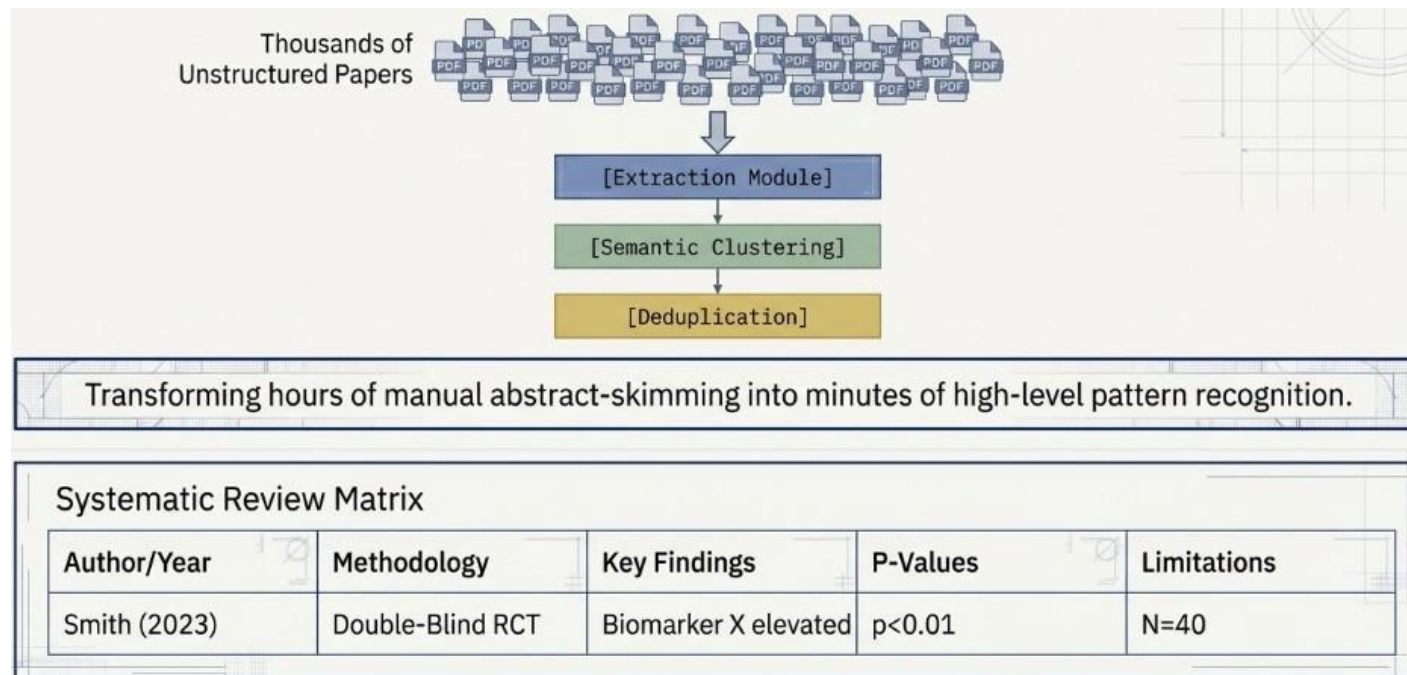
Mapping research landscapes

- AI can help identify dominant themes, methodological trends, research gaps, emerging topics

Human-in-the-loop requirement

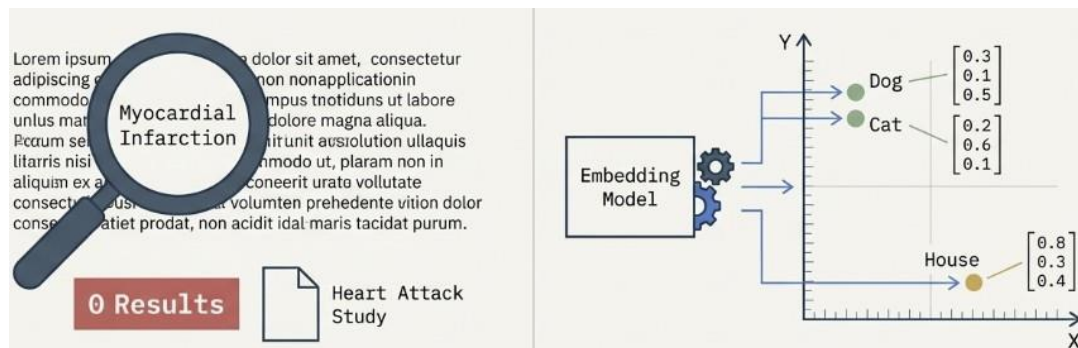
Risks and limitations

- hallucinated summaries, missed key papers, bias toward highly represented topics
- requires verification against original sources



AI enables literature reviews at a scale impossible to achieve manually, but human oversight remains essential for rigor, completeness, and critical interpretation

Semantic search vs keyword search in scientific databases



Keyword search (traditional)

- Matches exact words or phrases in titles, abstracts, and metadata
- Common in databases such as: PubMed, Scopus, Web of Science
- Strengths
 - precise and transparent
 - easy to reproduce
 - good for known terminology
- Limitations
 - misses synonyms and paraphrases
 - sensitive to spelling and wording
 - weak across interdisciplinary terminology

Semantic search (modern AI-based)

- Uses embeddings / vector representations
- Retrieves documents by conceptual similarity
- Example:
 - “AI bias in medicine”
 - may retrieve papers using terms like
 - “algorithmic fairness in clinical decision systems”
 - without exact keyword overlap.

Best practice: hybrid strategy

Keyword search retrieves exact terminology, while semantic search retrieves conceptual meaning; combining both improves completeness and rigor in scientific literature discovery

Summarizing complex scientific papers with LLMs

Motivation

- Scientific papers are dense, technical, time-consuming to read
- Challenge: Efficiently extracting key insights without losing critical detail

What LLMs can do

- Generate: concise summaries, structured overviews, simplified explanations
- Typical outputs: bullet-point summaries, section-wise breakdowns (methods, results, contributions)

Types of summaries

- Abstractive: rephrased, synthesized content
- Extractive: key sentences from the original text

Structuring scientific summaries

- Effective prompts can request research question, methodology, dataset and experiments, key findings, limitations

Multi-document summarization

- LLMs can synthesize across multiple papers: compare methods, identify trends, highlight gaps
- Useful for: literature reviews, meta-analysis preparation

Limitations

- risk of hallucinated details, omission of critical nuances
- difficulty with complex mathematical arguments, domain-specific precision

Best practices

- provide full text or relevant sections, structured prompts,
- verify summaries against original paper

LLMs can efficiently summarize complex scientific papers and synthesize knowledge, but their outputs must be carefully structured and critically verified to ensure accuracy and completeness

Citation-aware AI systems

Motivation: why citations matter

- traceability of claims
- verifiable sources
- standard LLMs often generate text without reliable citations
- Problem: fluent output without evidence undermines scientific rigor

What are citation-aware AI systems?

- AI systems that:
 - retrieve relevant documents
 - generate responses grounded in sources
 - provide explicit citations (e.g., page, paragraph)

How they work: RAG

Types of citation support:

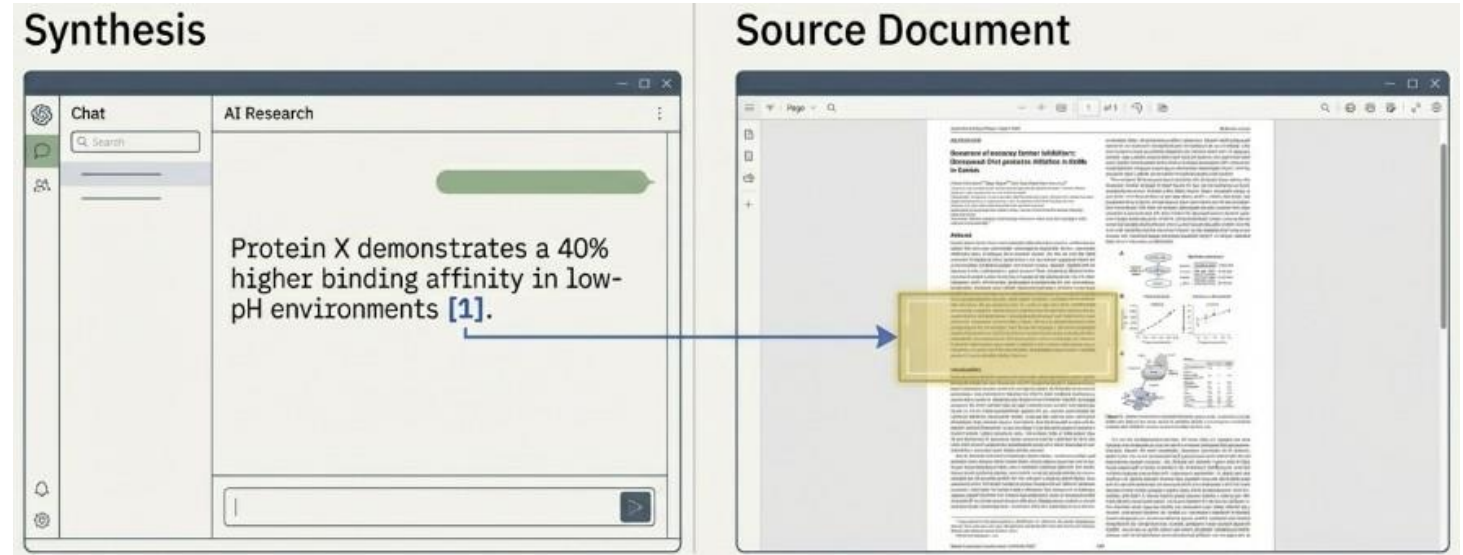
- document-level citation (paper title)
- passage-level citation (specific section)
- fine-grained citation (page/line references)

Benefits for research

Limitations

Evaluation Challenges

Research Implications



Citation-aware AI systems enhance trust in AI-generated content by grounding outputs in verifiable sources, but still require careful validation of citation accuracy and relevance

Why standard LLMs hallucinate

What is hallucination?

- Generation of factually incorrect, unsupported, or fabricated information
- Often appears fluent, coherent, and convincing

Root cause: probabilistic generation

Lack of grounding

Training data limitations

Overgeneralization

Instruction following vs truth

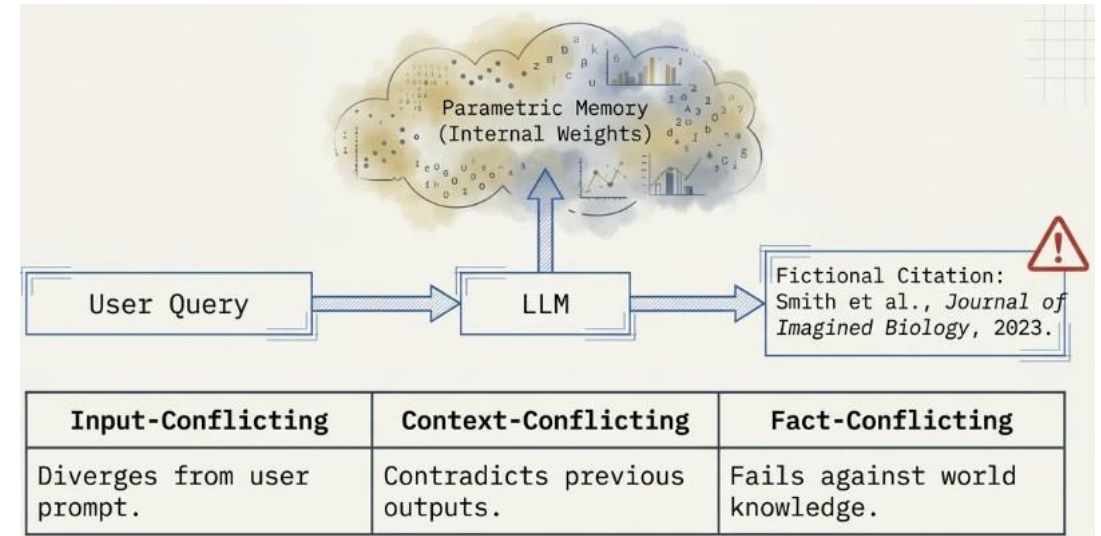
Lack of uncertainty awareness

Why this matters in research

- Risks include fabricated citations, incorrect summaries, misleading conclusions

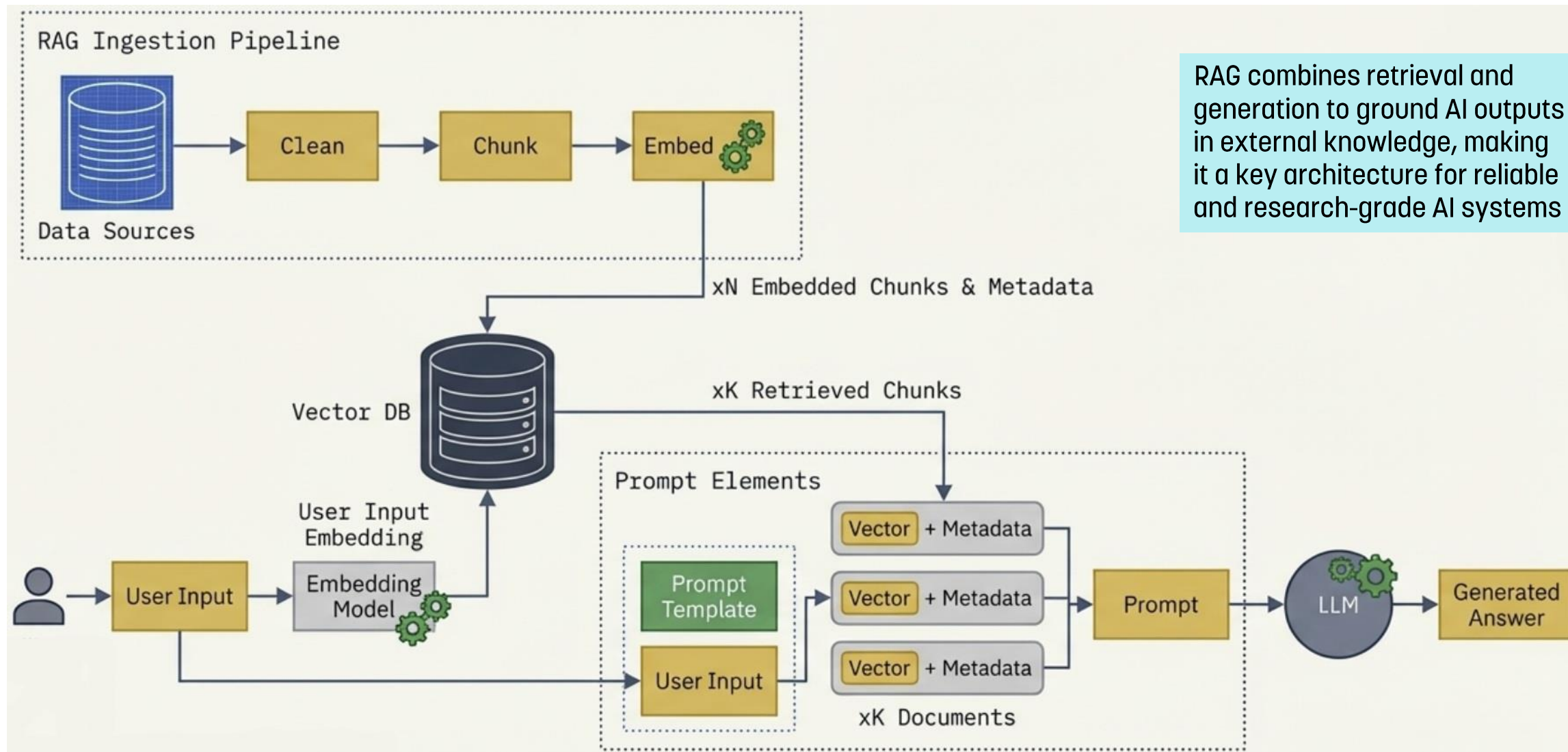
Mitigation strategies

- RAG
- citation-aware systems
- human verification



Standard LLMs hallucinate because they generate text based on statistical patterns rather than verified knowledge, making human oversight and grounding mechanisms essential in research use

Retrieval-Augmented Generation: concept and architecture



Embeddings, vector databases, and similarity search

From text to vectors

- text is converted into embeddings (dense vectors)
- embeddings capture semantic meaning, contextual similarity
- similar concepts are located close in vector space

What are embeddings?

- high-dimensional numerical representations of text
- learned from large corpora
- “cancer therapy” and “oncology treatment” → similar vectors

Vector databases

- specialized systems for storing and querying embeddings

Similarity search

- finds vectors closest to a query vector

Role in scientific research

- semantic literature search
- clustering research topics
- linking related studies
- finding relevant passages in large corpora

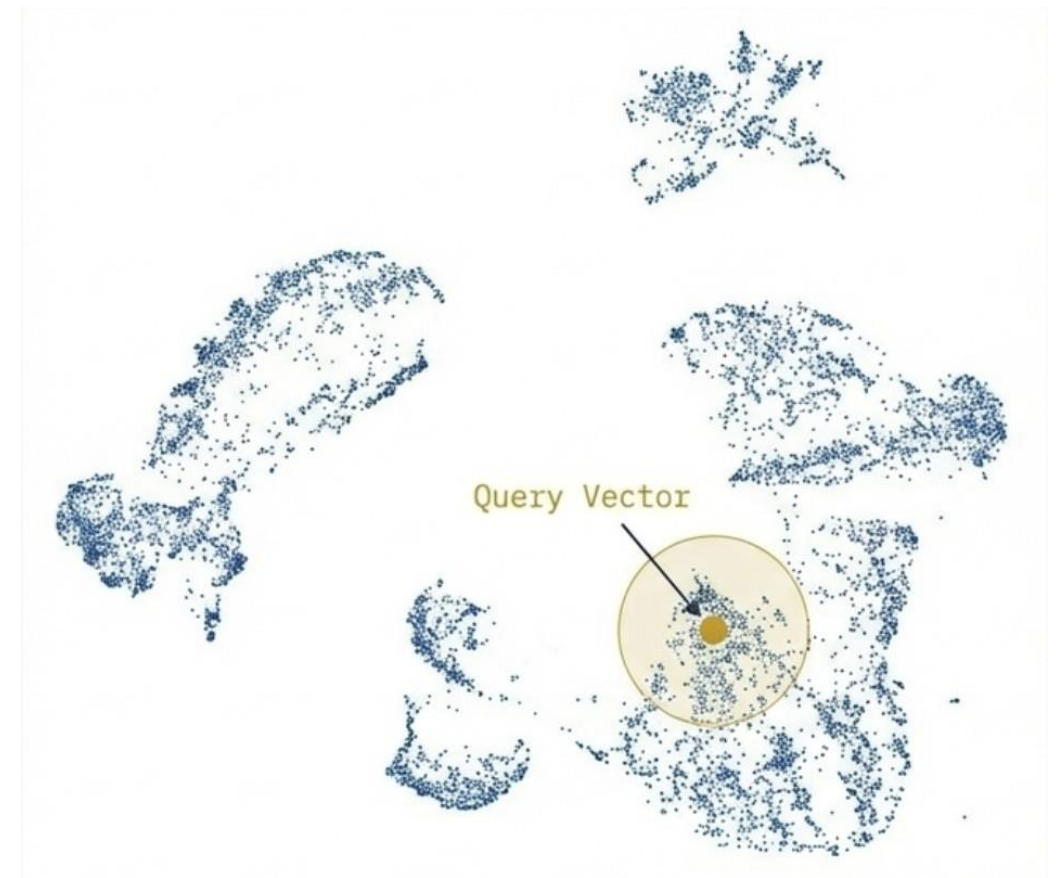
Integration in AI pipelines:

Text → Embedding → Vector DB → Similarity Search → Retrieved Context → LLM

Advantages

Limitations

Embeddings and vector databases enable semantic similarity search, forming the backbone of modern AI systems that retrieve meaning-based information rather than exact keyword matches



Chunking and context injection for scientific documents

Why chunking is necessary

- scientific papers are often longer than an LLM's context window
- full-text input may exceed token limits

What is chunking?

- Splitting documents into smaller text segments (“chunks”)
- Typical chunk sizes: paragraph-level, section-level, sliding windows (e.g., 300–1000 tokens)

Why chunk quality matters

Poor chunking can:

- separate definitions from explanations
- split tables from interpretation
- fragment methods and results

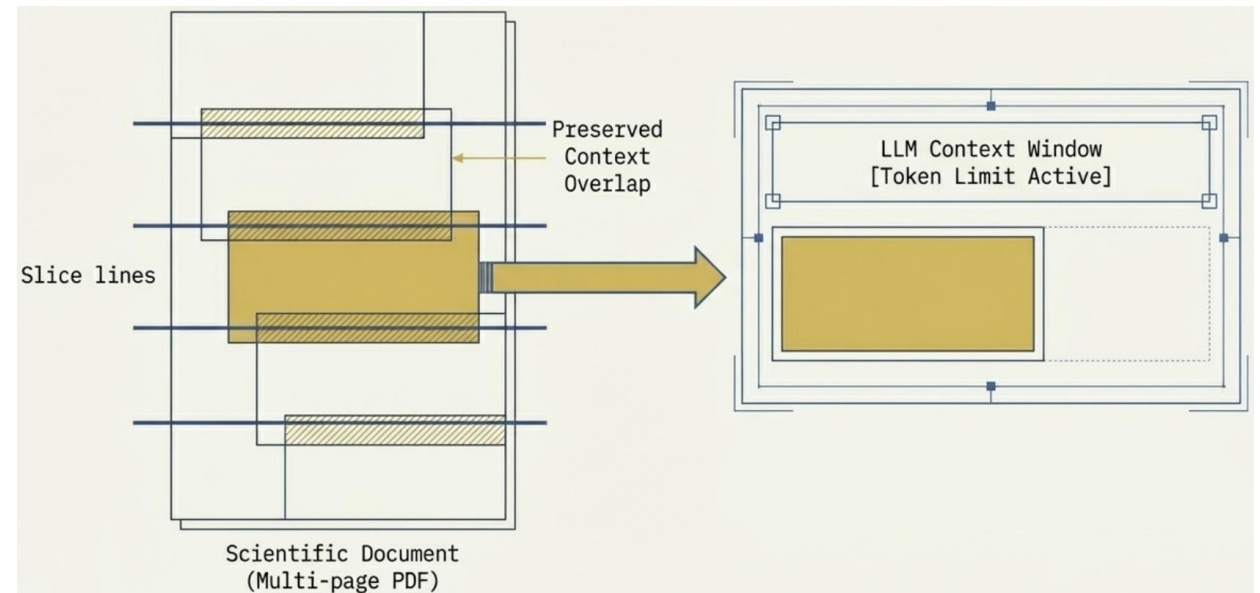
Context injection

- retrieved chunks are inserted into the prompt context
- LLM uses these chunks as grounded evidence

Scientific use cases

- Particularly useful for: long research papers, systematic reviews, protocols and technical reports

Role in RAG systems: retrieval quality, citation precision, hallucination reduction



Effective chunking and context injection are essential for reliable use of scientific documents in RAG systems, directly influencing retrieval precision, context quality, and output faithfulness

Evaluating and integrating RAG in research workflows

Why evaluation matters

- RAG systems may appear reliable because they cite sources. However, citations alone do not guarantee correctness.
- Does the system retrieve the right evidence and use it faithfully?

What to evaluate

- retrieval quality
- generation faithfulness
- end-to-end task performance

Human evaluation is essential

Integration into research workflow

- literature review pipelines
- evidence synthesis
- citation-aware writing
- question answering over paper collections

Risks of poor integration

- over-trust in retrieved content
- propagation of source errors
- omission of important, but unretrieved papers

Evaluating RAG requires assessing both retrieval quality and generation faithfulness, while successful integration places the system inside a human-verified research workflow rather than as an autonomous decision-maker

AI tools for knowledge management and research notes

Once grounded knowledge is retrieved, it must be organized into reusable research memory

Why knowledge management matters

- Research generates large volumes of papers, notes, annotations, code snippets, ideas and hypotheses
- Information must remain searchable, structured, and reusable

Role of AI tools

- automatic note summarization
- semantic organization of ideas
- linking related concepts across projects
- fast retrieval of prior notes

AI-Assisted note structuring

- method notes → assumptions, variables, limitations

Semantic search across notes

- Search by meaning, not exact wording
- Useful when researchers forget exact terms but remember the concept

Cross-Linking and Knowledge Graphs

Research Workflow Integration

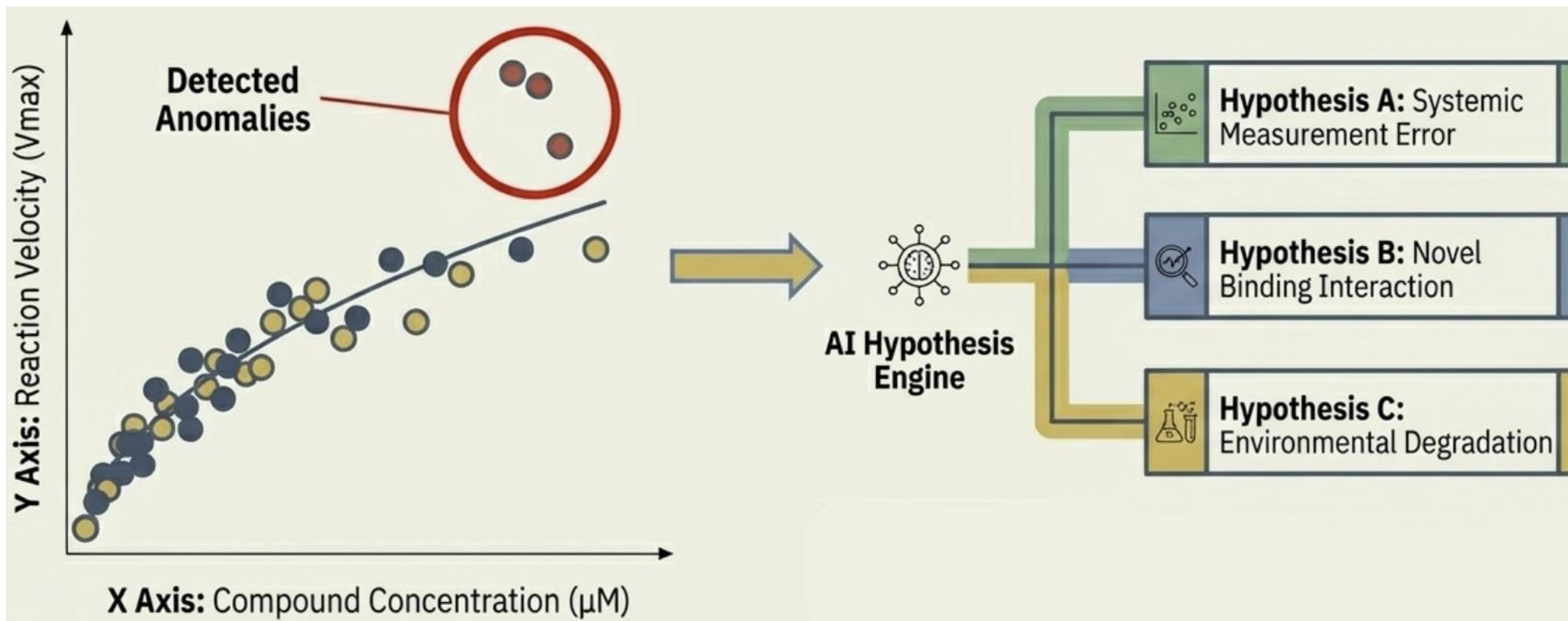
Common Tools / Approaches

- AI-enhanced notebooks
- semantic note systems
- literature knowledge bases

Risks and best practice: do not outsource interpretation entirely, preserve original notes, verify auto-generated summaries

AI transforms research notes from static documents into an intelligent, searchable knowledge system that supports long-term scientific thinking and reproducibility

AI-assisted data interpretation and hypothesis generation



AI can accelerate interpretation and hypothesis generation by identifying patterns and proposing plausible explanations, but scientific validity depends entirely on human theory, domain expertise, and empirical testing

AI for scientific writing: drafting, editing, and ethics

AI as a Writing Assistant

- LLMs support: drafting sections (abstract, introduction, related work), restructuring arguments, improving clarity and coherence
- ! AI accelerates writing, but does not replace authorship

Drafting Scientific Content

Editing and Refinement

Ethical Concerns

- authorship and attribution
- plagiarism and originality
- transparency of AI use
- accountability for errors

Guidelines and Policies

- Increasing requirements to disclose AI assistance, ensure human responsibility
- ! AI cannot be listed as an author

Risks

- hallucinated citations
- fabricated or distorted claims
- over-reliance on generated text

Best Practices

- verify all content against sources
- use AI for drafting and editing
- retain human control over arguments, conclusions

Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models

Tiffany H. Kung, Morgan Cheatham, ChatGPT, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, Victor Tseng

doi: <https://doi.org/10.1101/2022.12.19.22283643>

Now published in *PLOS Digital Health* doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)

“An attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs,” says Magdalena Skipper, editor-in-chief of Nature in London. Authors using LLMs in any way while developing a paper should document their use in the methods or acknowledgements sections, if appropriate, she says.

AI can significantly enhance scientific writing through drafting and editing, but ethical use requires transparency, verification, and full human responsibility for the final content

AI-assisted programming and LLM-based research systems

AI as a coding co-pilot

- code generation, debugging, refactoring, documentation

```
// Load genomic dataset, run PCA, and plot top 3 clusters  
pca = PCA(n_components=3); clusters = pca.fit_transform(data)
```

Research programming use cases

- data preprocessing pipelines, statistical analysis scripts, visualization code, simulation workflows, reproducible notebooks

Prompt-to-code interaction

- “Write Python code to perform PCA and visualize the first two components.”

Debugging and code explanation

- LLMs help with error interpretation, code review, explanation of unfamiliar libraries, optimization suggestions

Rapid prototyping

- quick testing of ideas, algorithms, data workflows

Risks and limitations

- syntactically correct but logically wrong code
- hidden bugs
- poor statistical assumptions
- insecure or inefficient implementations

Best practice

- Always verify correctness, reproducibility, statistical validity, software dependencies

AI-assisted programming accelerates research software development and experimentation, but researchers remain responsible for correctness, reproducibility, and scientific validity

- AI augments—not replaces—researchers
- grounding is essential for trust
- prompting is a methodological skill
- outputs require verification
- responsibility remains human

AI accelerates implementation, but scientific validity remains a human responsibility