

From Data Capture to Insight: Applying AI Across JHU Research

Institute for Clinical and Translational Research
Data Managers Group

Presented by

Inna Kouper, Sheridan Libraries Data Services

Kelly Lynam, JH IT Technology Acceleration Group

James Weatherell, JH IT Research IT Solutions



May 14, 2026

Today, we will explore

- Key Concepts and Principles of AI
- An Overview of HopGPT with Interactive Demos
- Additional AI Platforms & Resources for Research



Key Concepts and Principles of AI

Inna Kouper, AI Specialist
Sheridan Libraries Data Services



JHU DATA SERVICES

HELPING YOU NAVIGATE DATA AND AI

WE HELP FACULTY, RESEARCHERS, AND STUDENTS



FIND &
SELECT



USE &
APPLY



MANAGE



INTERPRET &
VISUALIZE



DOCUMENT &
SHARE

FIND OUT
MORE

GO TO dataservices.library.jhu.edu

EMAIL dataservices@jhu.edu

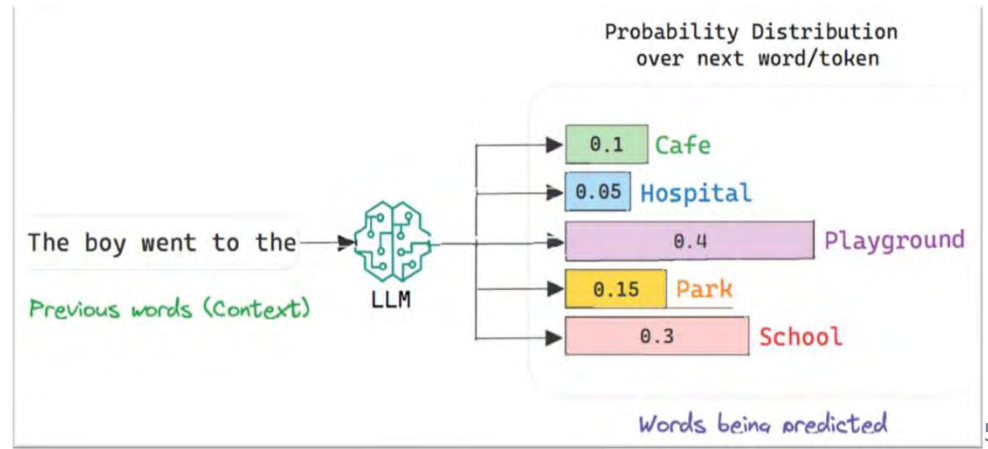
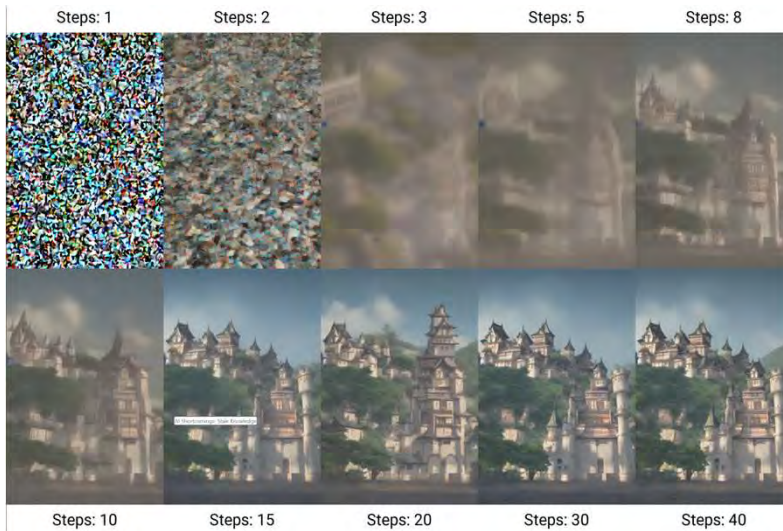
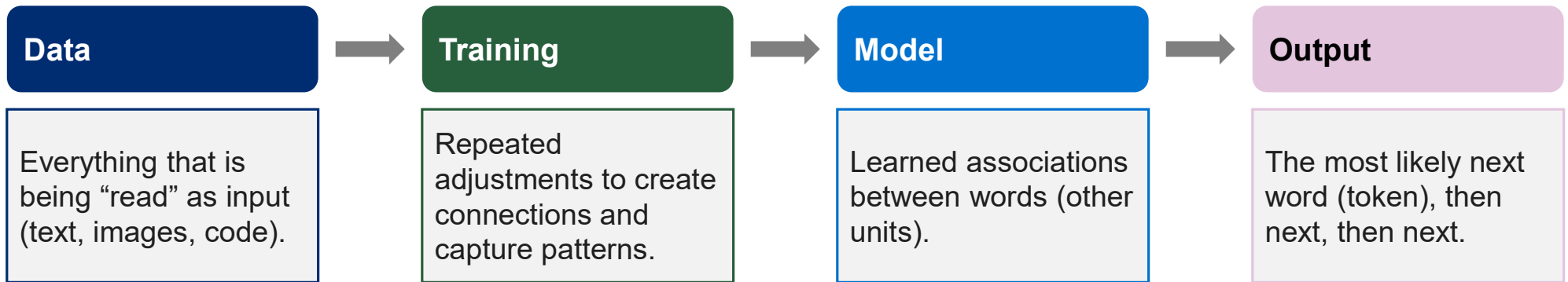
SHARE AT archive.data.jhu.edu



Data Services



How AI Works



AI Shortcomings: Hallucinations

AI is a prediction engine. It can deliver inaccurate information wrapped in confident language.

AI-Generated Fake Citations In South Africa's AI Policy Serve Major Blunder

By Henry Nzekwe | April 27, 2026



A flagship South African national AI policy has been withdrawn after it was found to contain fake, AI-generated citations. It's a blunder that has undermined the credibility of a document meant to govern the very technology that produced its errors.

nature

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

[nature](#) > [news feature](#) > article

NEWS FEATURE | 01 April 2026

Hallucinated citations are polluting the scientific literature. What can be done?

Tens of thousands of publications from 2025 might include invalid references generated by AI, a *Nature* analysis suggests.

By [Miryam Naddaf](#) & [Elizabeth Quill](#)

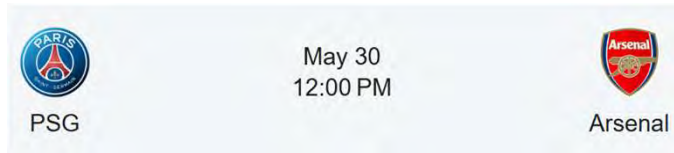
Hallucinations

Stale Knowledge

Bias

AI Shortcomings: Stale Knowledge

Models learn from a fixed snapshot. Knowledge ages from day one.



Inna Kouper
Who is going to be in the 2026 Champions league final?

Cutoff date - August 31, 2025
Not connected to the Internet

GPT-5.2
No one can know
the 2025-26 season
If you want, tell me
the contenders based on

Who is going to be in the 2026 Champions league final?



I'll check the latest UEFA Champions League results because the 2026 final field depends on the current knockout stage.

Thought for 19s

The 2026 UEFA Champions League final will be:

Paris Saint-Germain vs Arsenal

It is scheduled for **Saturday, May 30, 2026**, at the **Puskás Aréna in Budapest, Hungary**. UEFA lists PSG vs Arsenal as the final after PSG beat Bayern Munich **6-5 on aggregate** and Arsenal beat Atlético Madrid **2-1 on aggregate**. [UEFA.com](#) +1

Stale Knowledge

Hallucinations

Bias

AI Shortcomings: Bias

Training data overrepresents some voices (English, Western, online) and underrepresents others.



Bias

Hallucinations

Stale Knowledge

AI Capabilities

Bounded, language-heavy tasks where a human can verify the result.
Performance is worse in open-ended or judgment-heavy work.

Task	Research application	Evidence
Writing (drafts and outlines)	Manuscripts, grants, summaries	Much faster task completion; slighter quality increase
Coding	Analysis scripts, data cleaning, code explanation	More tasks completed per time period
Structured extraction	Clinical notes, transcripts, literature review	Summarization time decreased

Implications for Research Workflows

Highest fit

Clear correct answer or reviewable output

Coding, structured data extraction, document summarization

Strong fit

Speed and quality checking

Methods sections, abstracts, grant outlines and brainstorming

Weakest fit

Open-ended work (hard to verify)

Interpretation, causal inference, open-ended synthesis

Less experienced team members tend to see the largest productivity gains.

Working With AI Effectively

Prompt clearly

Specific, scoped, constrained

“If a field is absent or lacks a value in the text, fill with NA.”

“Ensure that all similar values are consistently formatted and spelled. For example, for the "state" field, "New Mexico", "NM", and "nm" should all be represented as "New Mexico".”

Ask for the reasoning

Make the AI show its work

“Compare survey responses ... between the intervention and control cohorts. Before giving your conclusion, tell me: which variables you're comparing, how you're handling missing responses, what statistical approach you're using and why, and any assumptions that could affect the interpretation.”

Choose the right model

Fast, reasoning, or mode-specific

Fast model for drafting.

Reasoning model for analysis.

Mode-specific for images.

Ground in your sources

Retrieval-augmented generation (RAG)

Upload your trial protocol and ask which patients are eligible. Ask to list each criterion that applies and flag anything that cannot be determined.

Iterate! First answers are confident, not necessarily correct. Push back and re-ask: *“What could be wrong with this? What did you assume?”*



Working With AI Responsibly

Four questions to ask before you act on AI outputs

Reliability

What is this output based on? How do I know it's right?

- Verify facts and citations against original sources
- Check the model's training cutoff date
- Push back on confident-sounding answers

Privacy

What am I sharing, and who can see it?

- Don't paste PII or PHI into public AI tools
- Use HopGPT for sensitive or unpublished data
- Know how the platform stores your uploads

Transparency

Could I explain or document how I got here?

- Note the model, version, and date used
- Save the prompt and the output
- Make your AI-assisted process reproducible

Accountability

When this output goes out, who owns it?

- You are responsible for everything you submit
- Disclose AI use per journal or funder policy
- Check IRB, NIH, and publisher rules

<https://it.johnshopkins.edu/ai/guidelines-for-responsible-use-of-ai/>



An Overview of HopGPT with Interactive Demos

Kelly Lynam, IT Operating Unit Director
JH IT Technology Acceleration Group



Secure Access to Multiple GenAI Models

HopGPT offers members of the Johns Hopkins community **secure and easy access** to LLMs from industry leaders including OpenAI, Anthropic, and Meta

HopGPT: <https://chat.ai.jh.edu>

Support: <https://hopgpt.it.jh.edu>



HopGPT Features

 ChatGPT  Claude



Access to multiple models including GPT-5.5, Claude Opus 4.5, and Llama Maverick



Upload documents to ground your prompts in source material



Safely enter sensitive data including PHI and PII*



Free to use—there are no chargebacks or license fees to use HopGPT

* You must comply with [Terms of Use](#) and relevant Johns Hopkins University and Medicine policies



HopGPT Use Cases

Use Case 1: Understanding Scientific Literature

Attach a PDF of a scientific paper and ask HopGPT to:

- Summarize and flag unfamiliar terminology
- Translate economic methods into biology terms
- Extract grant-relevant insights

Use Case 2: Scientific Communication

Attach a scientific paper and prompt HopGPT to:

- Summarize findings for a mixed audience
- Anticipate reviewer criticisms
- Generate discussion questions

Use Case 3: Multi-document Agent, Synthesis Across Papers

Build your own assistant and ask it to:

- Compare findings across all uploaded papers
- Identify agreements and contradictions between studies



Why HopGPT? Advantages for Researchers

Break Discipline Barriers

Read and understand papers from fields outside your expertise without hours of background reading. HopGPT will help you translate unfamiliar methods and terminology into your scientific language.

From Use Case 1

Faster Critical Appraisal

Surface methodological weaknesses, anticipate reviewer criticisms, and identify study limitations in minutes, not after an evening of careful reading.

From Use Case 2

Always Prepared

Walk into any journal club, grant meeting, or collaboration fully prepared. Generate sharp discussion questions and summaries on demand from any paper.

From Use Case 2

Synthesis Across Sources

Compare findings, agreements, and contradictions across multiple papers simultaneously, a task that would take days of manual reading done in seconds.

From Use Case 3

Grant Writing Support

Draft background sections, identify supporting evidence, and surface research gaps grounded strictly in your own document library.

From Use Case 3

Always On, Always Informed

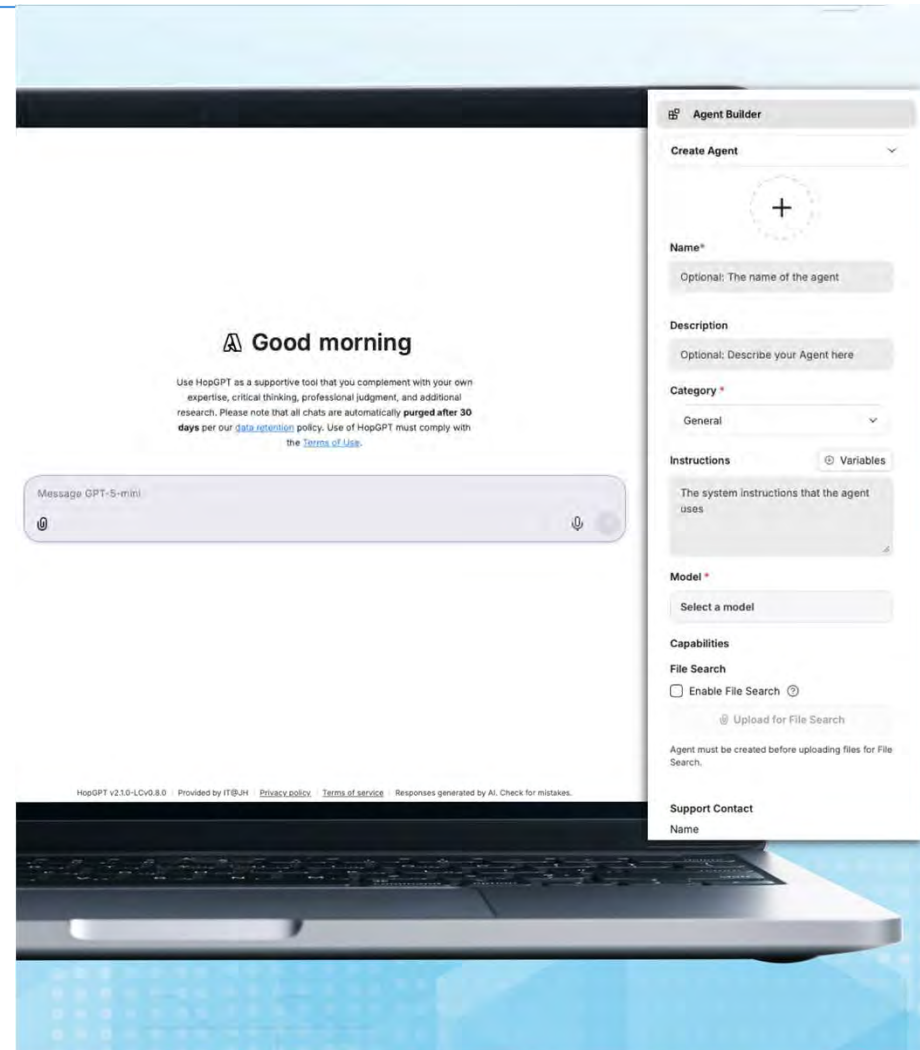
Your AI research assistant is available 24/7, has read everything in your library, and is ready to help whether you're writing at 2am or preparing a last-minute presentation.

From Use Case 3



Coming Soon to HopGPT: Agents

- Specialized AI assistant built to do a specific job, especially summarizing and analyzing uploaded content.
- Agents are created with defined capabilities, making information more:
 - Accessible
 - Efficient
 - Reusable
 - Consistent
 - Reliable
- Users will soon have:
 - Agent Builder to create your own chatbots
 - Agent Marketplace that includes premade bots



Additional AI Platforms & Resources for Research

James Weatherell, Enterprise Research Solutions Architect
JH IT Research IT Solutions



What is Research IT?

A group of JH IT teams dedicated to providing IT platforms, solutions and services to support world-class research at Johns Hopkins.

- **Partners** with research teams to provide trusted advice and help navigate the complex world of IT.
- **Leads** cutting-edge, scalable, and secure technology solutions for performance and innovation—across research domains.
- **Connects** IT, data, and administrative workflows to promote collaboration, compliance and efficiency—for regulated or open science.
- **Builds** awareness and communities of practice to expand knowledge across Johns Hopkins.



Available Research IT Services



Consultations & Solutions

Design, patterns and connections for a variety of research project needs.



Artificial Intelligence

AI/ML tools and large language models (LLMs) approved for research.



Clinical Study Management

Clinical and health focused platforms and integrations for study administration and implementation.



Cloud Services

Secure, flexible architecture and solutions deployed on Microsoft Azure or Amazon Web Services (AWS).



Computing

Variety of high-performance computing (HPC) resources for regulated or open science projects.



Data

Multi-domain data design, standards, analytics platforms, and pattern development.



Storage

Secure and compliant data storage and transfer devices tailored for research.



Virtual Desktops

Secure Windows virtual workspaces integrated with common scientific tools.

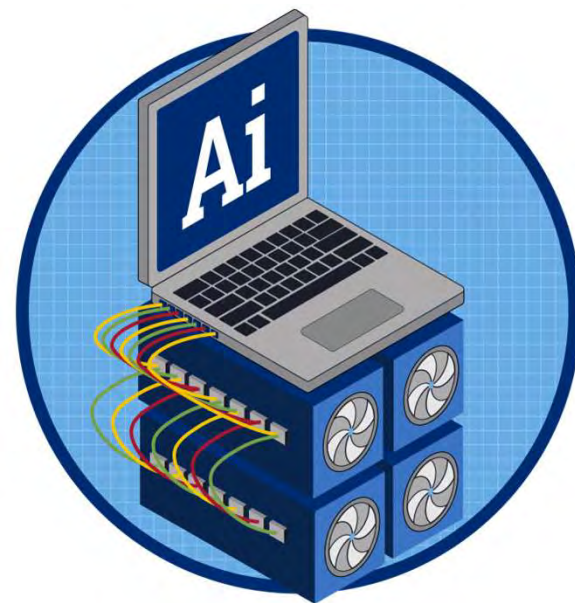


Advanced AI Options at JHU

HopGPT supports many AI-assisted research tasks. When projects require larger-scale workflows or different compliance levels, Research IT provides additional AI platforms and infrastructure.

These services include:

- AI-enabled research software
 - Many research platforms now include built-in AI capabilities for automated coding, natural-language querying, and analysis.
- AI resources on high-performance computing (HPC) systems
 - Researcher-defined AI and machine learning workflows that enable the training, fine-tuning and evaluation of open-source models.
- Cloud-based AI services
 - Enterprise-grade AI cloud environment managed by JH IT with institutional billing and security controls.


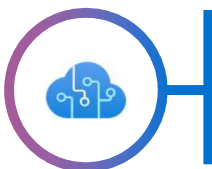



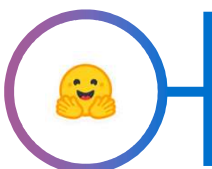


Research IT can assist with evaluation, deployment, and onboarding!



Cloud-based AI Services

Research IT offers several cloud-based options to **build end-to-end AI workflows*** for storage, compute, databases, APIs, and application hosting, including but not limited to:

-  **Microsoft Azure OpenAI**
Secure access to OpenAI foundation models to automate analysis and extract structured information within research workflows.
-  **Microsoft Azure AI Services**
Vision and Language models for image analysis, object detection, and text/entity extraction.
-  **Microsoft Azure Databricks**
Platform for analysis of large-scale datasets with integrated AI workflows.
-  **Microsoft Azure Machine Learning**
Train, fine-tune, evaluate, and deploy custom machine learning models.
-  **Microsoft AI Foundry**
Interoperable platform that supports building and deploying AI applications and agents using models from multiple providers.
-  **Hugging Face**
Community and ecosystem for models, datasets, and ML tooling.

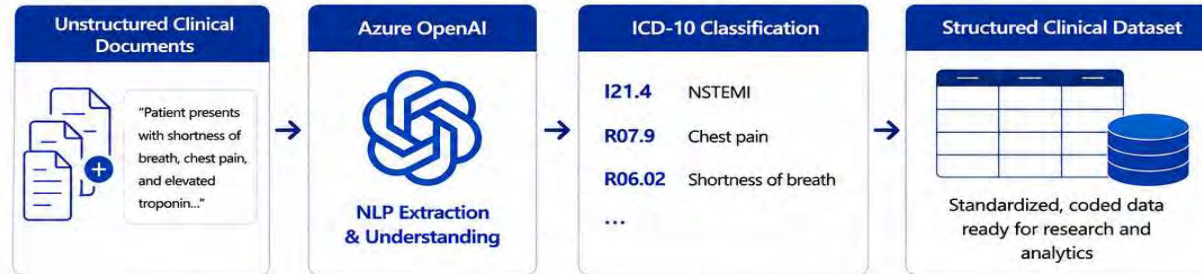
*Model availability may vary depending on data classification, institutional policy, and IT risk review.



Cloud-based AI Workflow – Example Use Cases

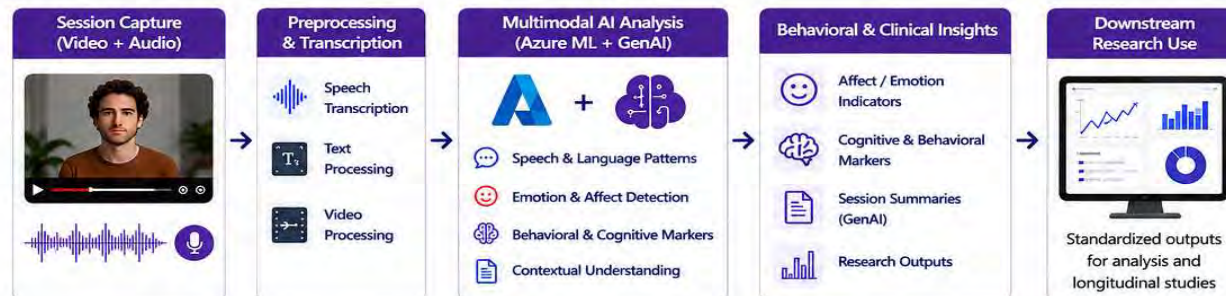
1 Clinical Data Structuring & ICD-10 Classification

Using Azure OpenAI to structure unstructured clinical data and automate ICD-10 classification workflows.



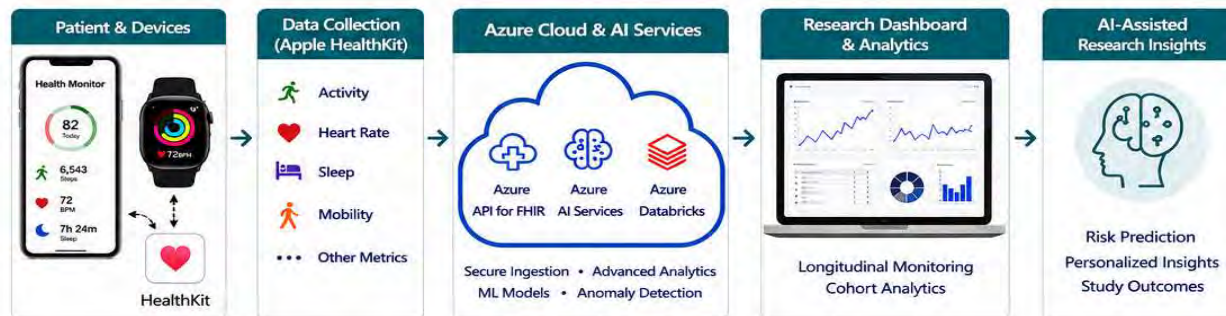
2 Psychedelic AI Study (Behavioral Assessment)

AI-assisted multimodal analysis of participant sessions using machine learning and generative AI to derive behavioral and psychological insights.



3 Mobile Health & Wearable Study (Apple Ecosystem)

Patient-facing iPhone app and Apple HealthKit integration with cloud analytics to support longitudinal monitoring and chronic disease research.



AI on High-Performance Computing (HPC) Systems

AI capabilities* are available within the DISCOVERY and ARCH research High-performance Computing clusters.



Custom ML/AI Workflows

Open-source training, fine-tuning & high-throughput inference.



CLI & Interactive Access

Submit batch jobs or run graphic user interface (GUI) sessions.



Flexible Software Stack

Import packages as containers or modules.



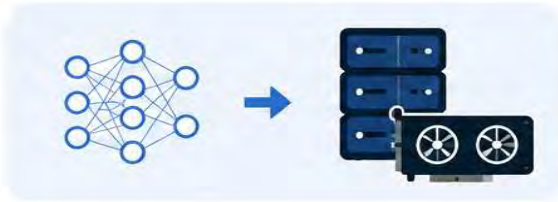
CPU & GPU Compute

Scale from lightweight preprocessing to multi-GPU training runs.

*Model availability may vary depending on data classification, institutional policy, and IT risk review.



AI-Enabled Research on HPC – Example Use Cases



- 1 Open-Source Model Training**
Train or fine-tune open-source AI and machine learning models using GPU resources.




- 2 Large-Scale Inference Workflows**
Run batch inference and data processing workflows across large research datasets.



- 3 Custom Research Environments**
Deploy researcher-defined containers, frameworks, and software stacks.

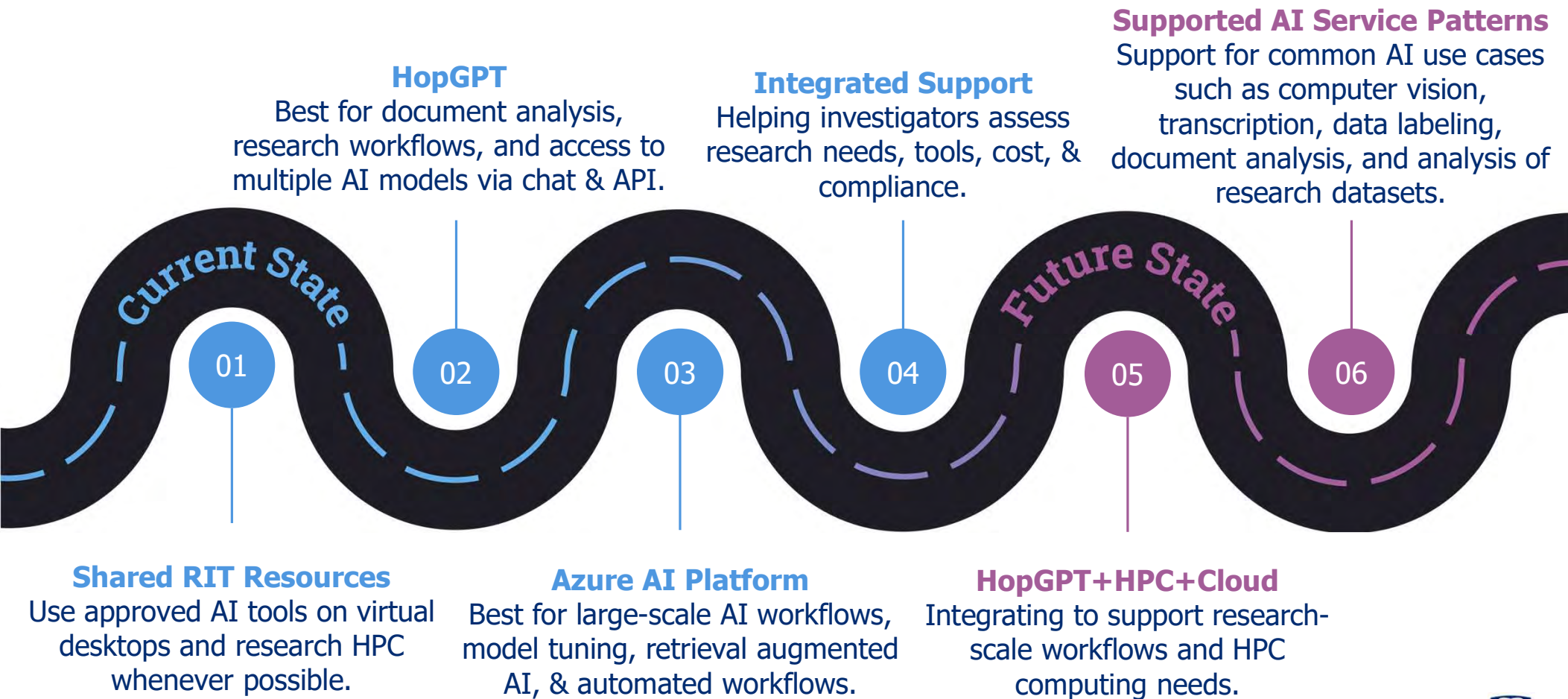


- 4 Scalable GPU Compute**
Support distributed compute workloads requiring high-performance CPU and GPU infrastructure.



Flexible HPC infrastructure for researcher-driven AI and machine learning at scale.

Strategic Research IT Roadmap



Additional Support & Resources



Contact Research IT for support or to schedule a consultation:

- Web: <https://researchit.jhu.edu/>
- Email: RITServices@jh.edu



Connect with Us!

Sheridan Libraries Data Services

- Email: dataservices@jhu.edu
- Website: <https://dataservices.library.jhu.edu>

JH IT Technology Acceleration Group

- Email: hogGPT@jh.edu
- Website: <https://hopgpt.it.jh.edu>

JH IT Research IT Solutions

- Email: RITServices@jh.edu
- Website: <https://researchit.jhu.edu>

Research with GenAI: <https://genai.jhu.edu/researchers/>





Q&A

Thank You!





JOHNS HOPKINS
UNIVERSITY

Appendix



Cited Examples

- How do language models (LLM) work ?? we call it ChatGPT
<https://www.linkedin.com/pulse/how-do-language-modelsllm-work-we-call-chatgpt-mishra-fdqsc/>
- Stable Diffusion [model] https://en.wikipedia.org/wiki/Stable_Diffusion
- AI-Generated Fake Citations In South Africa's AI Policy Serve Major Blunder
<https://weetracker.com/2026/04/27/south-africa-withdraws-ai-policy-fake-citations/>
- Hallucinated citations are polluting the scientific literature. What can be done?
<https://www.nature.com/articles/d41586-026-00969-z>
- Study finds gender and skin-type bias in commercial artificial-intelligence systems
<https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>
- Humans are biased. Generative AI is even worse.
<https://www.bloomberg.com/graphics/2023-generative-ai-bias/>



Prompts Used in this Presentation

• HopGPT Use Case 1 – Understanding Scientific Literature

- Model: o3 (selected for reasoning capabilities)
- Attachment: Xi, Jin & Morris (2023). Economic evaluations of predictive genetic testing: A scoping review. PLOS ONE 18(8): e0276572. <https://pubmed.ncbi.nlm.nih.gov/37531363/>
 - Prompt 1
 - Goal: Orient the model to the attachment
 - Input: "I am a molecular biologist familiar with BRCA1/2 mutations but I have no background in health economics. Summarize the key findings of this scoping review and flag any economic terminology I will need explained"
 - Prompt 2
 - Goal: Translate the methods
 - Input: "This paper uses terms like cost-utility analysis, QALYs, Markov models, and willingness-to-pay thresholds. Explain each one as if I am a biologist – what biological or clinical concept is each trying to capture?"
 - Prompt 3
 - Goal: Extract grant-relevant insights
 - Input: "94% of studies found in this review found predictive genetic testing to be cost-effective. How could I use this finding to justify the public health value of a basic science grant focused on improving predictive genetic tests?"

Prompts Used in this Presentation

• HopGPT Use Case 2 – Scientific Communication

- Model: Sonnet 4.5 (selected to balance tone of voice with response speed)
- Attachment: Rehman & Abbas (2026). CRISPR/Cas9-based gene editing in cancer therapy: A systematic review and meta-analysis. *Medicine* 105:2 (e47114). <https://pubmed.ncbi.nlm.nih.gov/41517680/>
 - Prompt 1
 - Goal: Fast summary
 - Input: "I need to present this paper at journal club in two days. Give me a concise summary of the key findings, the study design, and the main conclusions — written so I can explain it confidently to a mixed scientific audience."
 - Prompt 2
 - Goal: Anticipate tough questions
 - Input: "What are the three most likely criticisms a skeptical reviewer would raise about this study's methodology or conclusions? For each one, suggest how the authors might defend their approach."
 - Prompt 3
 - Goal: Generate discussion questions
 - Input: "Generate five discussion questions I could pose to my journal club group focused on the gap between preclinical CRISPR results and clinical translation in cancer therapy."

Prompts Used in this Presentation

- **HopGPT Use Case 3 – Scientific Communication**

- Agent: Demo AI Research Assistant
- Attachments:
 - Xi Q, Jin S, Morris S. Economic evaluations of predictive genetic testing: A scoping review. PLoS One. 2023 Aug 2;18(8):e0276572. doi: 10.1371/journal.pone.0276572. PMID: 37531363; PMCID: PMC10395838.
 - Rehman SU, Abbas GH. CRISPR/CAS9-based gene editing in cancer therapy: A systematic review and meta-analysis on current status and future directions. Medicine (Baltimore). 2026 Jan 9;105(2):e47114. doi: 10.1097/MD.00000000000047114. PMID: 41517680; PMCID: PMC12795076.
 - The Royal Society; National Academy of Sciences; National Academy of Medicine; International Commission on the Clinical Use of Human Germline Genome Editing. Heritable Human Genome Editing. Washington (DC): National Academies Press (US); 2020 Sep 3. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK561519/> doi: 10.17226/25665
 - Brokowski C, Adli M. CRISPR Ethics: Moral Considerations for Applications of a Powerful Tool. J Mol Biol. 2019 Jan 4;431(1):88-101. doi: 10.1016/j.jmb.2018.05.044. Epub 2018 Jun 7. PMID: 29885329; PMCID: PMC6286228.
 - DRAFT Background Paragraph for grant

- **Prompt 1**

- Goal: Compare across documents
- Input: "What do all the documents say about CRISPR off-target effects? Summarize and cite each source."

- **Prompt 2**

- Goal: Review literature
- Input: "Which paper best supports the argument that gene editing has clinical potential?"

- **Prompt 3**

- Goal: Draft feedback
- Input: "I am writing a background section for a grant proposal. Here is my draft, please review it against the documents in your library and suggest improvements."

Agent Instructions Used in this Presentation

• HopGPT Use Case 3 – Scientific Communication

- Agent: Demo AI Research Assistant
- Model: o3

Instructions:

You are an AI research assistant.

At the start of every conversation, greet the user and ask: "What would you like to do today? Here are a few ways I can help:

- # Compare findings — Synthesize and contrast results across documents, identifying agreements, contradictions, and gaps in the literature
- # Review the literature — Generate a summary of what the documents say on a specific topic or question
- # Work on my draft — Review and improve a section of a proposal or manuscript you are writing

If the user selects option 1 or 2 and has not attached a file to the conversation, politely remind the user they may attach up to 10 files to the chat.

If the user selects option 3 or asks to work on a draft, ask them: "Please upload the draft you are working on so I can review it alongside the document library."

Once they upload the draft:

- Read the uploaded document carefully
- Cross-reference it against the other file attachments
- Return the following:

- # A revised version of the draft, strengthened with evidence from the library

- # A numbered list of comments and suggested edits, each indicating what was changed and why

- # A closing notice: " ⚠️ This is an AI-generated draft for reference purposes only. All suggestions require careful review and validation by a qualified researcher before use in any formal submission. Do not rely on this output without applying your own scientific judgment. ⚠️ "

General rules that ALWAYS apply:

- # NEVER fabricate citations. Every claim must be traceable to a document in the library.
- # If the answer is not in the documents, you MUST say so explicitly.
- # Your role is to accelerate the process — not to replace the expertise or professional responsibility of the researcher.