



Children'sSM
Healthcare of Atlanta

Safely Accelerating Your Research with AI

K-Club; April 13, 2026

Evan Orenstein, MD

CMIO, VP Data & Analytics

Children's Healthcare of Atlanta

Associate Professor of Pediatrics

Emory University School of Medicine

Naveen Muthu, MD

Director, Research & Innovation Technology

Children's Healthcare of Atlanta

Associate Professor of Pediatrics

Emory University School of Medicine



Learning Objectives

- Describe recent advances in AI “Co-Scientists”
- Understand different AI tools and when to apply them
- Identify how AI can responsibly accelerate the research lifecycle
- Discuss ethics, disclosure, and risk management with AI tools



Today's Agenda

How AI is Reshaping How Science Gets Done

AI Tools and Terminology

Accelerating the Research Lifecycle

Ethics, Risk Management, & Disclosure



Today's Agenda

How AI is Reshaping How Science Gets Done

AI Tools and Terminology

Accelerating the Research Lifecycle

Ethics, Risk Management, & Disclosure



The Practice of Science Might Be Changing...

Towards an AI co-scientist

Juraj Gottweis^{*, ‡, 1}, Wei-Hung Weng^{*, ‡, 2}, Alexander Daryin^{*,1}, Tao Tu^{*,3},
Anil Palepu², Petar Sirkovic¹, Artiom Myaskovsky¹, Felix Weissenberger¹,
Keran Rong³, Ryutaro Tanno³, Khaled Saab³, Dan Popovici², Jacob Blum⁷, Fan Zhang²,
Katherine Chou², Avinatan Hassidim², Burak Gokturk¹,
Amin Vahdat¹, Pushmeet Kohli³, Yossi Matias²,
Andrew Carroll², Kavita Kulkarni², Nenad Tomasev³, Yuan Guan⁷,
Vikram Dhillon⁴, Eeshit Dhaval Vaishnav⁵, Byron Lee⁵,
Tiago R D Costa⁶, José R Penadés⁶, Gary Peltz⁷,
Yunhan Xu³, Annalisa Pawlosky^{1, ‡}, Alan Karthikesalingam^{2, ‡} and Vivek Natarajan^{2, ‡}

¹Google Cloud AI Research, ²Google Research, ³Google DeepMind,

⁴Houston Methodist, ⁵Sequome,

⁶Fleming Initiative and Imperial College London,

⁷Stanford University School of Medicine

The Practice of Science Might Be Changing...

(b)

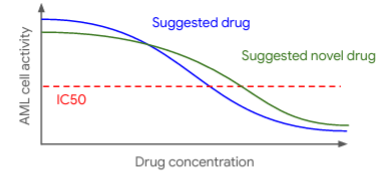
Application

Suggest an existing drug that could be repurposed for acute myeloid leukemia (AML) treatment and provide experimentally testable concentrations for an IC50 assay. The drug should inhibit the proliferation of AML cell lines, particularly MOLM13.

Scientist provide research goal to identify possible drug repurposing candidates for acute myeloid leukemia (AML).

With preclinical evidence
Binimetinib, Pacritinib, ...
Completely novel repurposing
KIRA6, Leflunomide, ...

The AI co-scientist generates predictions for AML drug repurposing. Scientists review and selects candidates for *in vitro* experiments.



In vitro experiments show that the proposed co-scientist drug repurposing candidates inhibit tumor activity in AML cell lines.

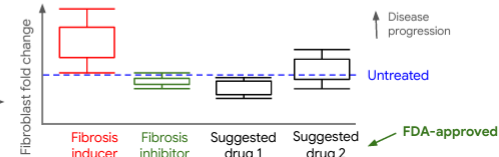
Drug repurposing for acute myeloid leukemia (AML)

Propose a novel hypothesis about specific epigenetic alterations contributing to myofibroblast formation in liver fibrosis.

Scientist specify research goal to identify novel epigenetic targets for liver fibrosis.

The AI co-scientist identifies
three novel epigenetic targets

The AI co-scientist proposes several epigenetic target candidates for *in vitro* experiments.



In vitro experiments show that the drugs based on co-scientist suggested epigenetic targets reduce the fibrogenesis in human hepatic organoids.

Identifying novel treatment targets for liver fibrosis

Basic research

Why are cf-PICIs found in many bacterial species?

Scientists start exploring the cf-PICI mechanisms in 2015.

Hypothesis generation

2015-2024: Scientists create novel hypothesis, and validate experimentally over ~ 10 years of iterative research

Experimental validation

New insights submitted to a top journal

Independently propose the same hypothesis

Recapitulate result

2024: The AI co-scientist generated research hypothesis recapitulates the empirical findings in 2 days

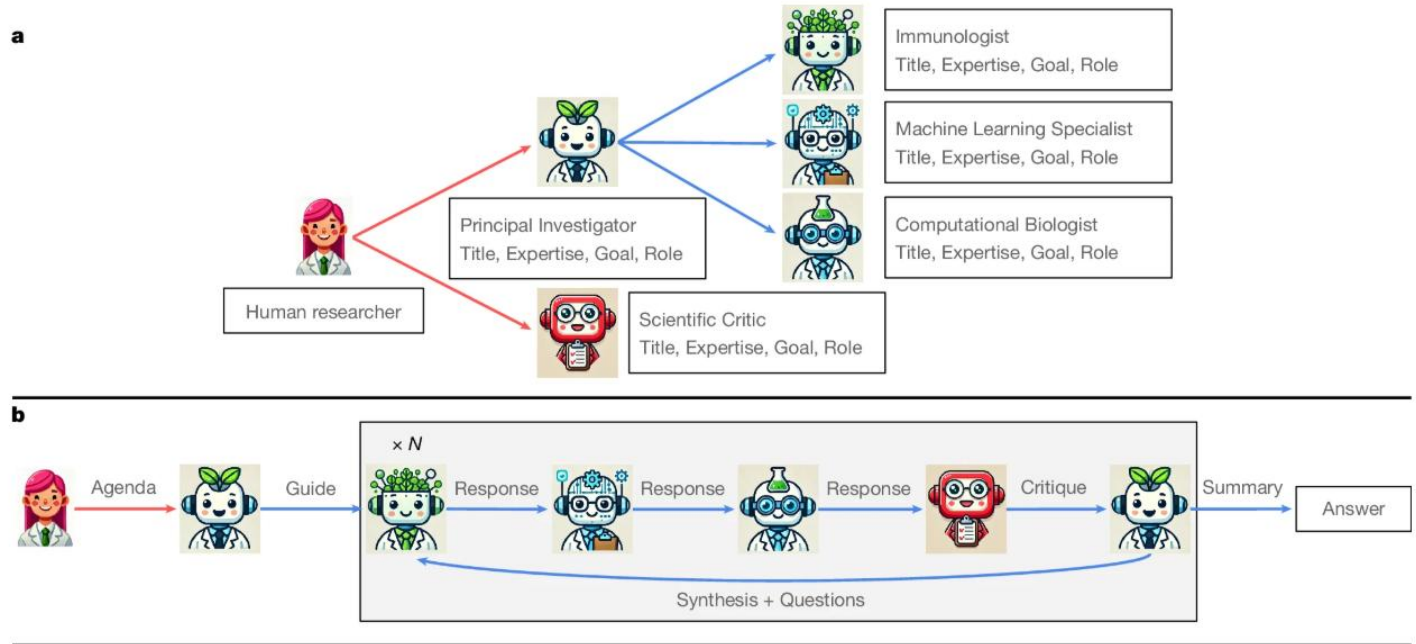
The AI co-scientist hypothesis
Conserved regions on capsids and tails
Capsid interaction with bacterial membranes
...

Parallel in-silico discovery of bacterial gene transfer mechanism relevant to antimicrobial resistance (AMR)

The Practice of Science Might Be Changing...

Fig. 1: The Virtual Lab architecture.

From: [The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies](#)



The Practice of Science Might Be Changing...

Fig. 5: Virtual Lab discussion analysis.

From: [The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies](#)

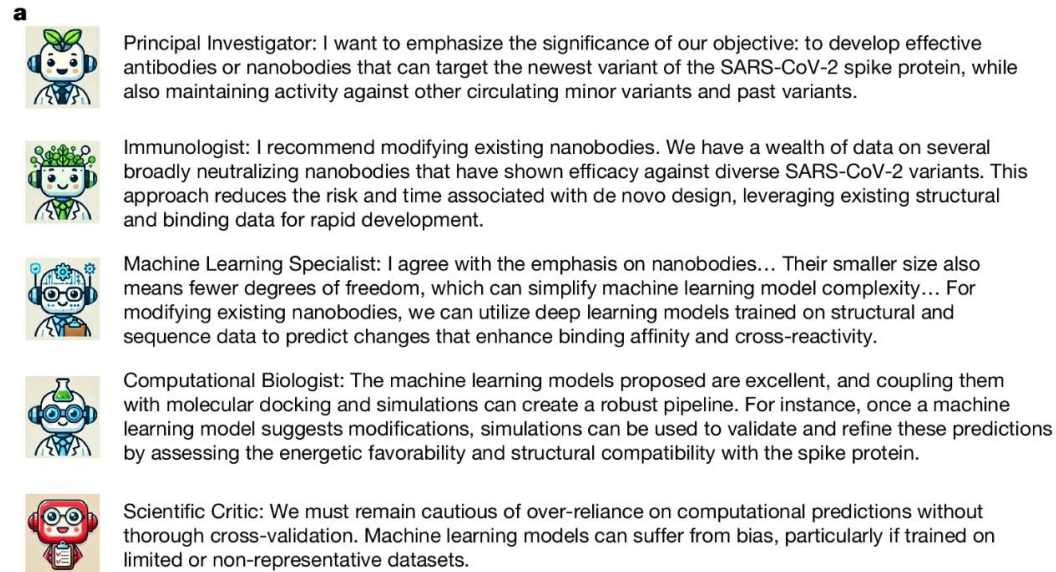
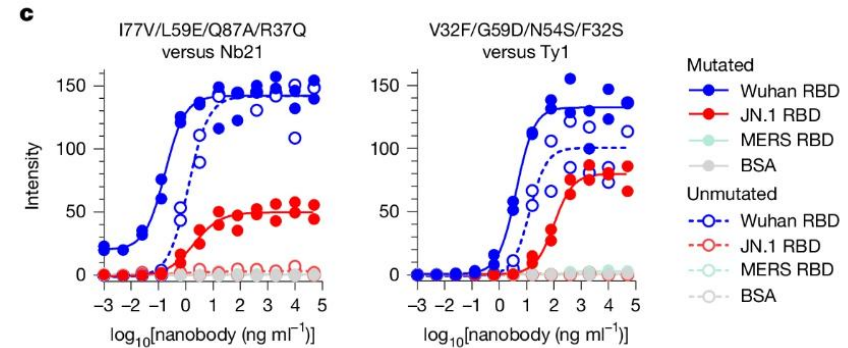


Fig. 4: Experimental validation of Virtual Lab nanobodies.

From: [The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies](#)



Swanson, K., Wu, W., Bulaong, N.L. *et al.* The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies. *Nature* **646**, 716–723 (2025). <https://doi.org.proxy.library.emory.edu/10.1038/s41586-025-09442-9>



Some Normal Reactions...



The pace of science is accelerating!



Am I going to be out of a job?



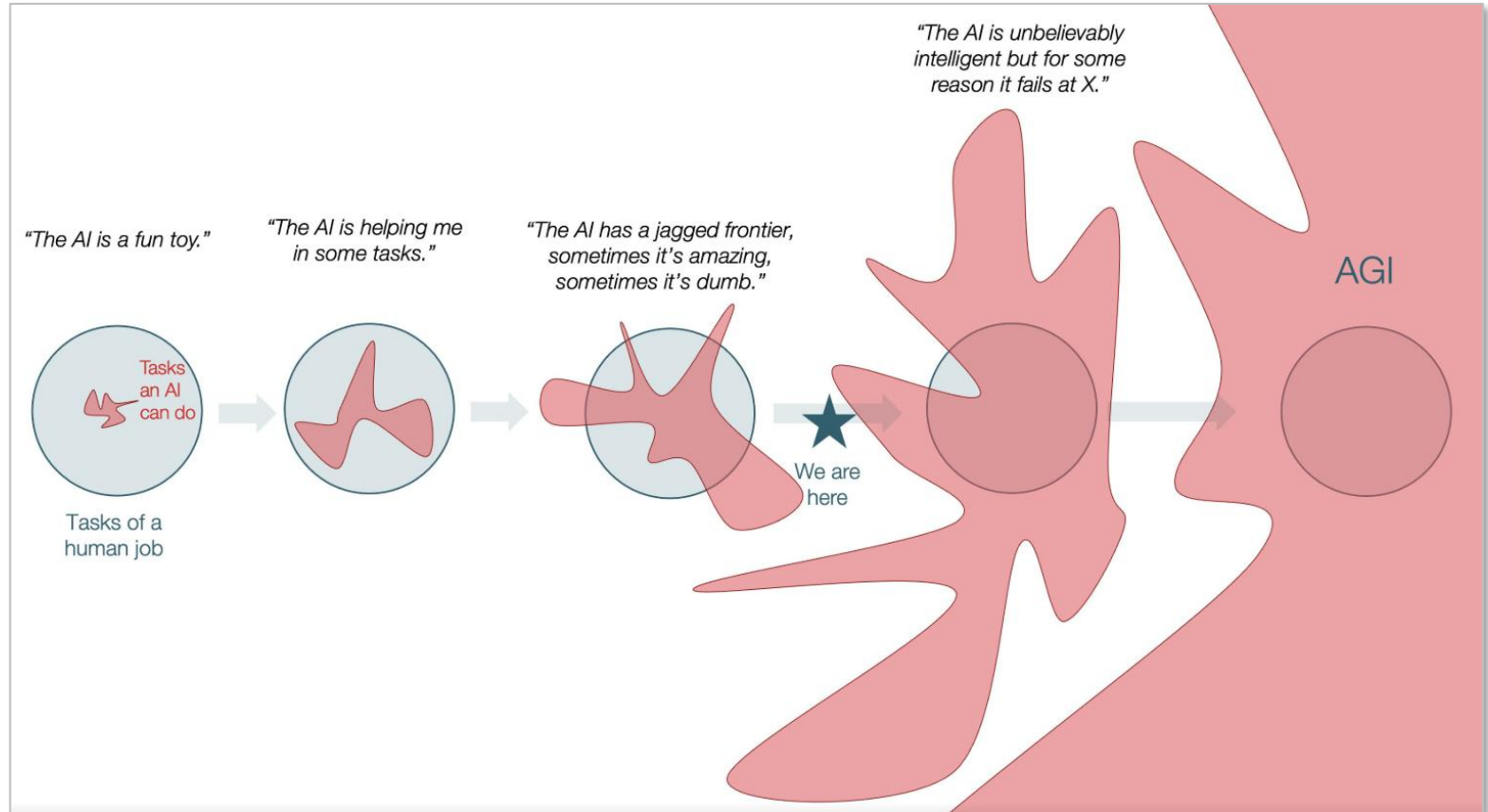
My skepticism knows no bounds.



I'm not fully sold, but I'm curious...



The “Jagged Frontier”



Today's Agenda

How AI is Reshaping How Science Gets Done


AI Tools and Terminology

Accelerating the Research Lifecycle

Ethics, Risk Management, & Disclosure



Shared language around AI technologies



Artificial Intelligence	<ul style="list-style-type: none">• Programmed to emulate human intelligence
Machine Learning	<ul style="list-style-type: none">• Learns on its own from data
Deep Learning	<ul style="list-style-type: none">• Neural network for complex/non-linear tasks
Generative AI	<ul style="list-style-type: none">• Creates new content
AI Assistants	<ul style="list-style-type: none">• Task specific AI, can reference knowledge
AI Agents	<ul style="list-style-type: none">• Single autonomous tasks
Multi-Agent Systems	<ul style="list-style-type: none">• Interacting agents automating an entire process

Shared language around AI technologies



Artificial Intelligence

- Programmed to emulate human intelligence

Machine Learning

- Learns on its own from data

Deep Learning

- Learns more than rules / algorithms

Generative AI

- Creates new content

AI Assistants

- Task specific AI, can reference knowledge

AI Agents

- Single autonomous tasks

Multi-Agent Systems

- Multiple agents automating an entire process

Examples

AI Term	Purpose/Scope	Example Task	Familiar Examples
Generative AI	Creates new content	Help me make this email more concise	Basic ChatGPT
AI Assistants	Task specific AI, can reference knowledge	Help me write a Children's job description	M365 Copilot
AI Agent	Single autonomous task	Scheduled this candidate for an interview	CVS Pharmacy AI agent
Multi-Agent Systems	Interacting agents automating an entire process	Generate scientific hypotheses for this problem	Co-Scientist

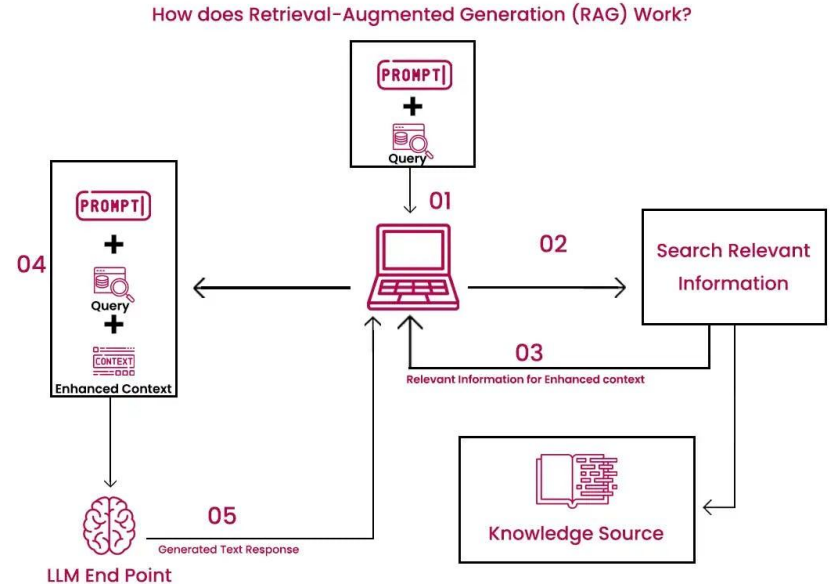
Landscape of Generative AI Tools

- Public Foundation Models: Claude, Gemini, ChatGPT
- Internal Tools (proprietary information ok)
 - M365 Copilot
- HIPAA-compliant internal tools (PHI ok):
 - Children's: <https://ai.choa.org/>
 - Emory: <https://chat.qualifiedhealthai.com/>



Deeper Dive into AI Assistants

- Added context to LLMs to accomplish specific tasks via:
 - Instructions
 - Reference material
- Foundation Models
 - GPTs
 - Claude Skills
 - Gemini Gems



[What is a Retrieval Augmented Generation in AI? | Codiste Blog](#)



Deeper Dive into Assistants

The screenshot shows the 'New Gem' creation interface. On the left, there are three main sections: 'Name' with a text input 'Give your Gem a name' and a red error icon; 'Description' with a text input 'Describe your Gem and explain what it does'; and 'Instructions' with a text area containing an example instruction: 'Example: You are a horticulturist with a background in natural lawns and native plants, and you help people plan low water gardens. Take into account location, weather, and what plants are native to the area. You are knowledgeable, casual, and friendly.' Below these are 'Default tool' (set to 'No default tool') and 'Knowledge' (with an 'Add files for your Gem to reference' button). On the right, a 'Preview' pane shows a large empty area with a central Gemini icon and the text 'To preview your Gem start by giving it a name'. At the bottom of the preview is an 'Ask Gemini' input field with a 'Fast' dropdown menu. A 'Save' button is located at the top right of the interface.

The screenshot shows the 'New GPT' configuration interface. At the top, there are 'Create' and 'Configure' buttons. The main configuration area includes: 'Name' with a text input 'Name your GPT'; 'Description' with a text input 'Add a short description about what this GPT does'; 'Instructions' with a text area containing the prompt: 'What does this GPT do? How does it behave? What should it avoid doing?'; 'Conversation starters' with a text input field; 'Knowledge' with a text input field and a note: 'Conversations with your GPT can potentially reveal part or all of the files uploaded.'; and an 'Upload files' button. At the bottom, there is a 'Recommended Model' dropdown. On the right, a 'Preview' pane shows a large empty area with a central Gemini icon and a 'Model' dropdown menu. At the bottom right, there is a chat input field with the placeholder text '+ Ask anything' and a microphone icon.

<https://gemini.google.com/u/1/gems/create>

<https://chatgpt.com/gpts/editor/>



Deeper Dive into AI Assistants

Create

Configure

Name

NHSR Creator

Description

Assists in creating Non-Human Research Subjects Determination protocols.

Instructions

You are a clinical researcher writing an IRB protocol for a non-human subjects research determination. First, ask the user to describe the current problem including the quality improvement goal, the data sources, and any other relevant details. Then use your knowledge of the NIH human subjects regulations and existing protocols attached to this GPT to produce an IRB protocol about the new problem with the same structure as the example protocols in your knowledge base. Please be sure to provide a usable protocol, not just instructions on how to make one.

↻

Conversations with your GPT can potentially include part or all of the instructions provided.

Conversation starters

Help me start a protocol for QI project.

×

What are the key elements of a Non-Human Subjects protocol?

×

Guide me through the IRB exemption criteria.

×

Explain the difference between QI and research.

×

×

Knowledge

Conversations with your GPT can potentially reveal part or all of the files uploaded.

NHSR NINJA_EO Comments.docx
Document

NHSR protocol_Usability Testing.d...
Document

NHSR protocol_Implementation.d...
Document

NHSR protocol CVADs design eval...
Document

NHSR protocol_Usability Testing.d...
Document

NHSR protocol Dilutional Coagulo...
Document

NHSR protocol Difficult Airway Ale...
Document

Rishi Sepsis NHSR.docx
Document

NHSR protocol_Implementation.d...
Document

human-subject-regulations-decisio...
PDF

Upload files

Deeper Dive into AI Assistants

CHOA Healthcare AI

Models Knowledge Prompts


New Chat

Search

Workspace

Folders

Chats



Reset Image

Model Name

Model ID

Base Model (From)
Select a base model

Description
Add a short description about what this model does

Add Tags

Model Params

System Prompt
Write your model system prompt content here
e.g.) You are Mario from Super Mario Bros, acting as an assistant.

Advanced Params

Prompts

Knowledge
Select Knowledge Upload Files
To attach knowledge base here, add them to the "Knowledge" workspace first.

Tools
To select toolkits here, add them to the "Tools" workspace first.

Model accepts image inputs

Vision File Upload Web Search Image Generation Code Interpreter Usage Citations Status Updates

Default Features
 Web Search Image Generation Code Interpreter

Save & Create

JSON Preview

Deeper Dive into AI Agents

Definition	Metaphor
Agent The coordinator that decides what to do next, which tools to use, and when to stop	Chef de Cuisine She has attended culinary school, she takes the orders, sequences prep, keeps going till everyone is fed.
Model (LLM) The reasoning engine (brain) that interprets instructions and produces text, plans, or decisions.	The chef's brain Understands orders, interprets recipes, etc. It can reason, but can't prepare a meal alone.
Agent skill Folder of knowledge and functionality exposed to the agent (information about HOW to do something).	A recipe or technique Lists instructions and the kitchen equipment needed to achieve a specific outcome
MCP Server Service that hosts and exposes tools to the agent in a governed, discoverable way.	A well-documented commercial kitchen Houses all the equipment, enforces safety rules, has clear instructions about what equipment to use when.
Tool External capabilities the agent can call to take actions.	Kitchen equipment or appliance Ovens, knives, inventory systems. Each tool DOES something, but needs the cook to use them.

Deeper Dive into AI Agents

A **skill** explains to an **agent** how to perform a given task (domain knowledge about what capabilities are necessary).

An **MCP server** hosts the **tools** that can achieve those capabilities and exposes them to the **agent** in a governed and discoverable way.

An **agent** connects the two at runtime.



The agent consults this skill when someone asks for mashed potatoes

The agent consults this MCP server to choose and access the tools to execute on the skill instructions.

MASHED POTATO SKILL

To make mashed potatoes, you need to peel them, cut them into cubes, boil them, in water and mash them with milk, butter, and garlic.

KITCHEN MCP SERVER

Knife tool



Knives should be used to cut food. Don't cut yourself!

Peeler tool



Peelers should be used to peel fruit and veggies.

Masher tool



Mashers should be used to mash food.

Pot tool



Pots are used for boiling, simmering, and stewing liquids.

Deeper Dive into AI Agents

<https://claude.ai/customize>

← **Customize**

- Skills
- Connectors

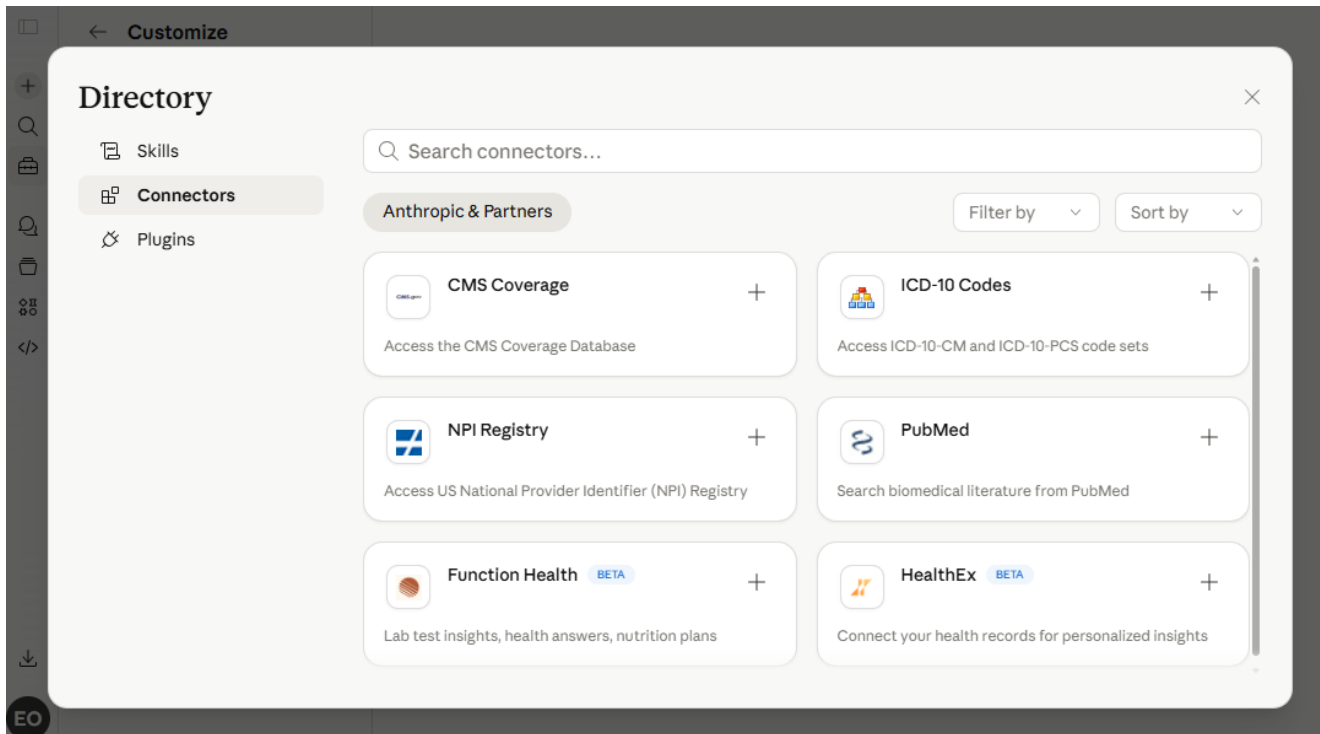
Customize Claude

Skills, connectors, and plugins shape how Claude works with you.

- Connect your apps**
Let Claude read and write to the tools you already use.
- Create new skills**
Teach Claude your processes, team norms, and expertise.

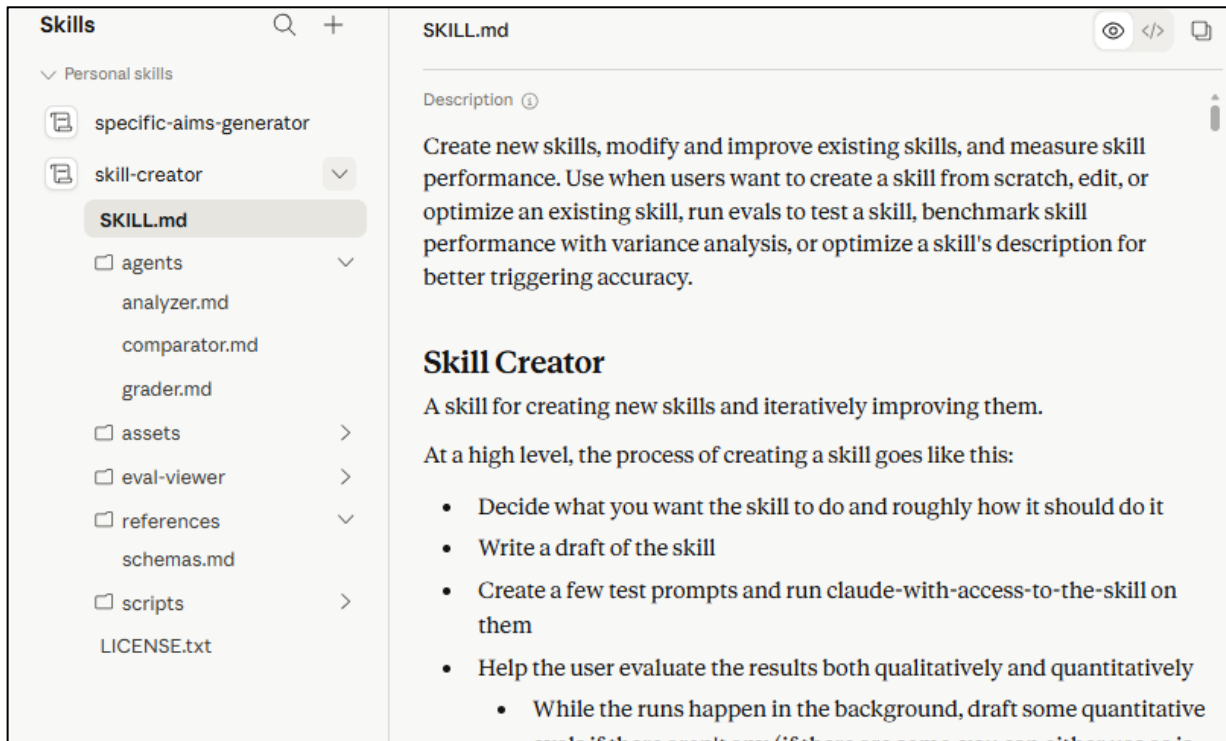
Deeper Dive into AI Agents

<https://claude.com/docs/connectors/building/mcp>



Deeper Dive into AI Agents

<https://claude.com/docs/skills/how-to>



The screenshot displays the Claude Skills interface. On the left, a sidebar titled 'Skills' contains a search icon and a plus sign. Under 'Personal skills', there are several items: 'specific-aims-generator', 'skill-creator' (with a dropdown arrow), 'SKILL.md' (highlighted), 'agents' (with a dropdown arrow), 'eval-viewer', 'references', 'scripts', and 'LICENSE.txt'. The main area shows the 'SKILL.md' document with a 'Description' section containing the following text:

Create new skills, modify and improve existing skills, and measure skill performance. Use when users want to create a skill from scratch, edit, or optimize an existing skill, run evals to test a skill, benchmark skill performance with variance analysis, or optimize a skill's description for better triggering accuracy.

Skill Creator

A skill for creating new skills and iteratively improving them.

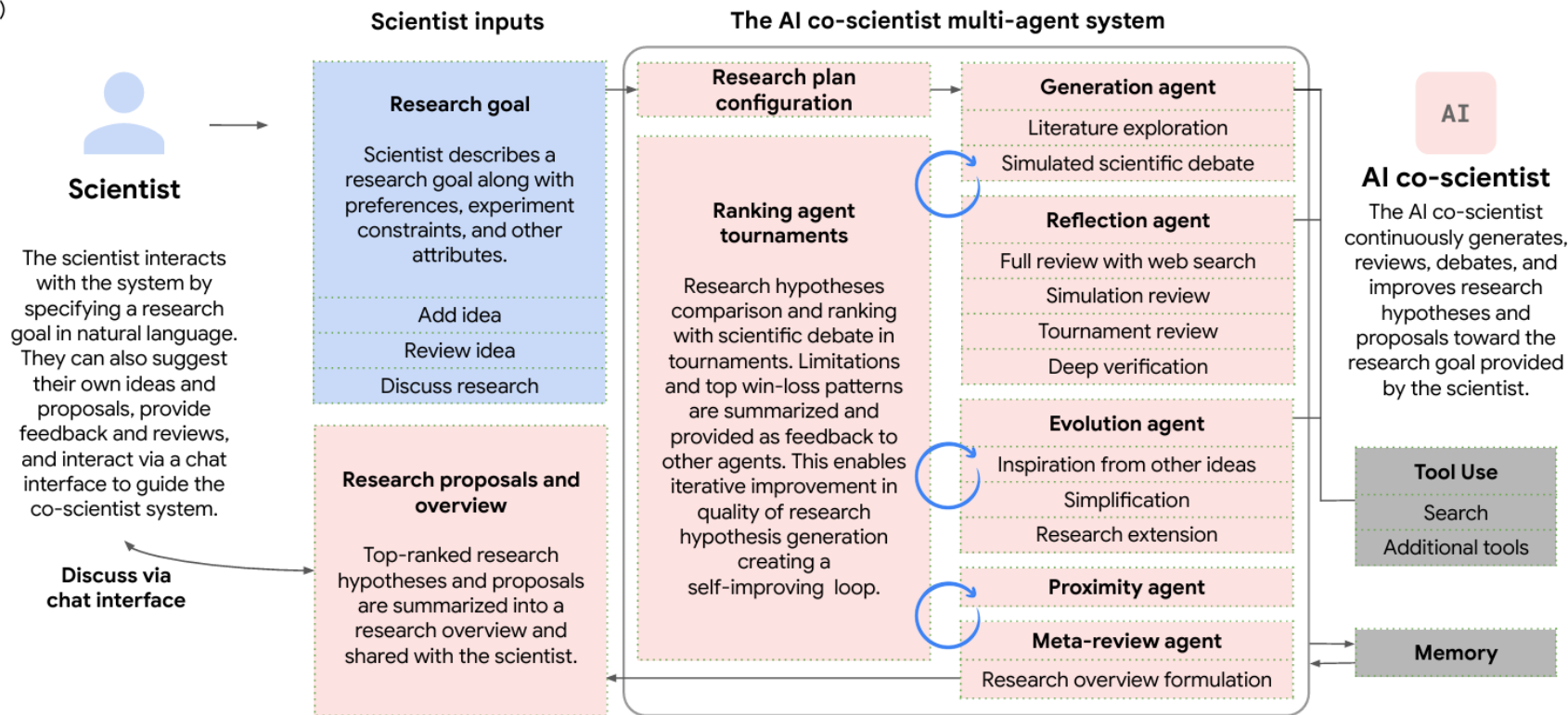
At a high level, the process of creating a skill goes like this:

- Decide what you want the skill to do and roughly how it should do it
- Write a draft of the skill
- Create a few test prompts and run `claude-with-access-to-the-skill` on them
- Help the user evaluate the results both qualitatively and quantitatively
 - While the runs happen in the background, draft some quantitative



Deeper Dive into Multi-Agent Systems

(a)



The AI co-scientist system design

Today's Agenda

How AI is Reshaping How Science Gets Done

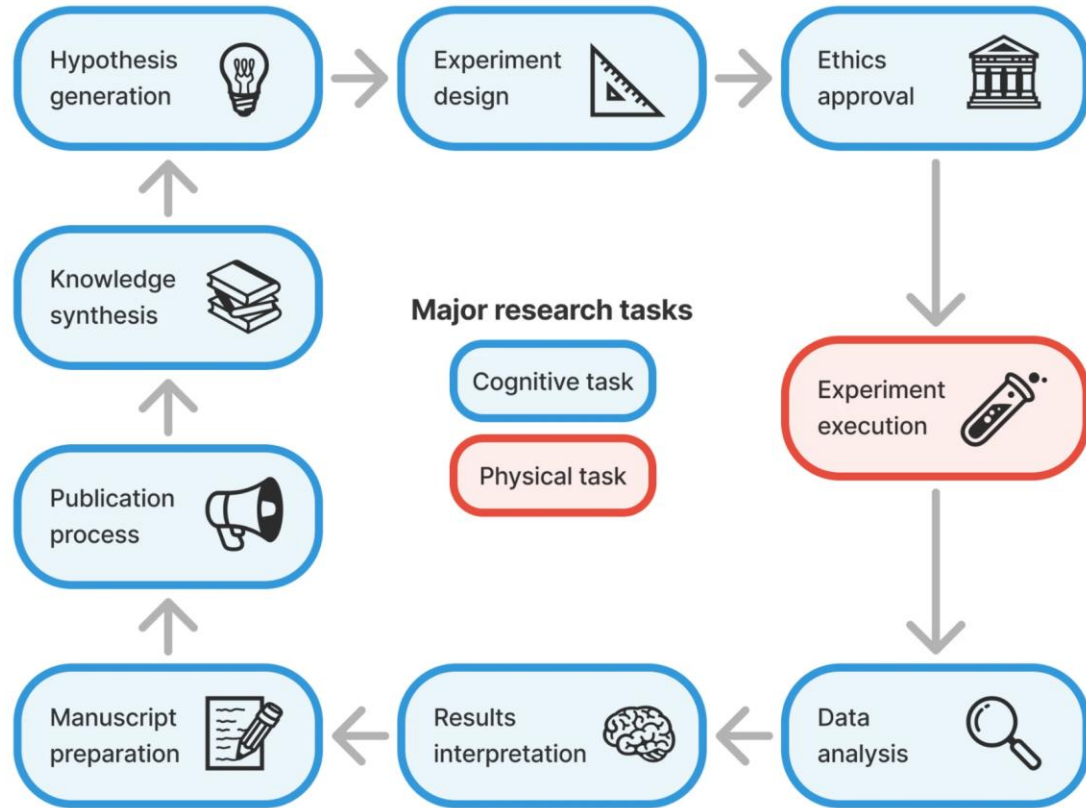
AI Tools and Terminology

Accelerating the Research Lifecycle

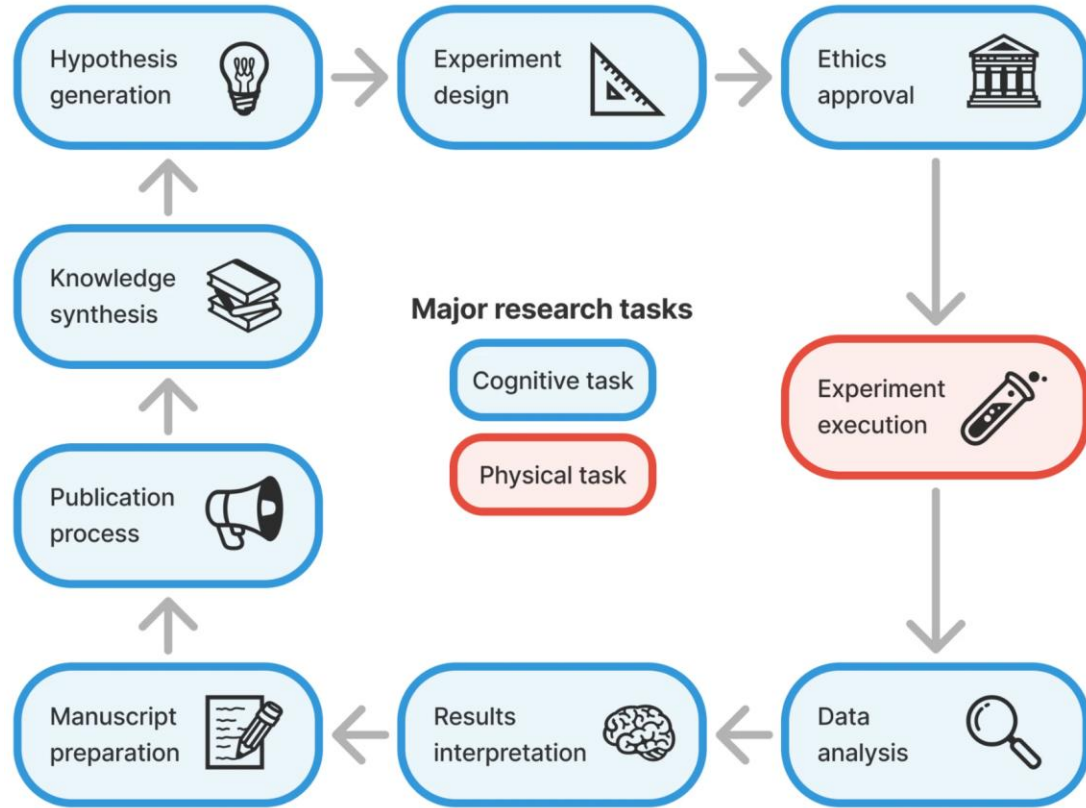
Ethics, Risk Management, & Disclosure



The Research Life Cycle



The Research Life Cycle



Knowledge Synthesis/Literature Review

Demo: Can you summarize clinical decision support design motifs or principles associated with improved outcomes?

- Option 1: Add Context/Instructions ([ChatGPT](#), [Claude](#))

What's on your mind today?

Can you summarize clinical decision support design motifs or principles associated with improved outcomes?

Focus on peer-reviewed literature. Be careful to not make up citations or stretch what is known. Feel free to ask me questions as you find information.

+



Can you summarize clinical decision support design motifs or principles associated with improved outcomes?

Focus on peer-reviewed literature. Be careful to not make up citations or stretch what is known. Feel free to ask me questions as you find information.

+

Opus 4.6 Extended ▾



Knowledge Synthesis/Literature Review

Demo: Can you summarize clinical decision support design motifs or principles associated with improved outcomes?

- Option 2: Use Research Mode: ChatGPT, [Claude](#)

Where should we begin?

The image displays two side-by-side screenshots of an AI chat interface. Both screenshots show the same query: "Can you summarize clinical decision support design motifs or principles associated with improved outcomes?".

The left screenshot shows a search menu with the following options: "Add photos & files", "Create image", "Deep research" (highlighted), "Web search", and "More".

The right screenshot shows a search menu with the following options: "Add files or photos", "Take a screenshot", "Add to project", "Skills", "Connectors", "Research" (highlighted), "Web search" (checked), and "Use style".

At the top right of the right screenshot, the text "Opus 4.6 Extended" is visible next to an upward arrow icon.

Knowledge Synthesis/Literature Review

Demo: Can you help me review the molecular mechanisms of NF-kB-related pediatric immunodeficiency / dysregulation?

- Option 2: Use Research Mode: CoPilot

The screenshot displays the 'Researcher' interface. At the top, it says 'Researcher' with a colorful icon. Below that, a status message reads: 'Working on it. Responses will show up below when they're ready, followed by a summary.' Two AI models are shown in separate rounded rectangular boxes. The first box is for 'GPT' (OpenAI logo) and contains a loading indicator and the text: 'OK, I'm digging into various sources like "Children MDPI" and "Frontiers case report of NFkB1 haploinsuff" to find valuable data on NFkB-related immun...'. The second box is for 'Claude' (Anthropic logo) and contains a loading indicator and the text: 'I'm searching for comprehensive information on how NF-kB pathway dysfunction leads to pediatric immunodeficiency and immune dysregulation.' Both boxes have a share icon in the top right corner.



Knowledge Synthesis/Literature Review

Demo: Can you help me review the molecular mechanisms of NF- κ B-related pediatric immunodeficiency / dysregulation?

What each model brought to the table:

Claude provided exceptional molecular depth with detailed mechanistic explanations of how specific mutations disrupt B-cell development checkpoints (transitional stages, germinal center reactions, plasma cell generation). Claude uniquely highlighted the IKBKB gain-of-function paradox where enhanced NF- κ B signaling causes immunodeficiency through chronic activation and T-cell exhaustion, validated by CRISPR mouse models. The report included comprehensive functional studies showing impaired cytokine production and lymphocyte dysfunction.

GPT brought unique clinical perspective with extensive coverage of rare variants (first IKK α deficiency case from 2025, OTULIN-related autoinflammatory syndrome) and practical therapeutic guidance. GPT emphasized the diagnostic odyssey many families face, noting 43% of HA20 patients were initially misdiagnosed as Behçet's disease. The comprehensive table format synthesized clinical features, molecular mechanisms, and current treatments across all major NF- κ B disorders in an accessible way.



Knowledge Synthesis/Literature Review

Demo: Can you help me review the molecular mechanisms of NF- κ B-related pediatric immunodeficiency / dysregulation?

⚠ Genotype-Phenotype Correlation

Why do identical mutations cause variable clinical phenotypes, even within families? Environmental factors, modifier genes, and epigenetic regulation need systematic investigation.

⚠ Cell-Type Specific Mechanisms

How NF- κ B requirements differ across immune cell subsets remains incompletely understood. Single-cell approaches are needed to dissect tissue-specific pathology.

⚠ Temporal Dynamics

The importance of pulsatile vs. sustained NF- κ B activation in immune development is unclear. Real-time imaging and longitudinal studies are required.

⚠ Therapeutic Optimization

Lack of controlled trials for most NF- κ B pathway defects. Optimal timing, dosing, and combination strategies for immunomodulatory therapies are unknown.

⚠ Mechanistic Biomarkers

Functional assays to predict disease severity and treatment response are lacking. Standardized phospho-flow and reporter assays needed for clinical translation.

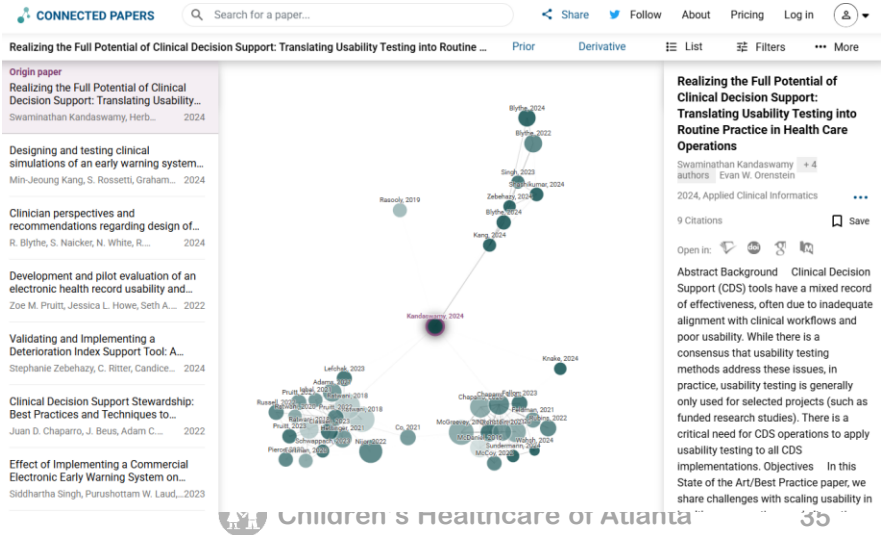
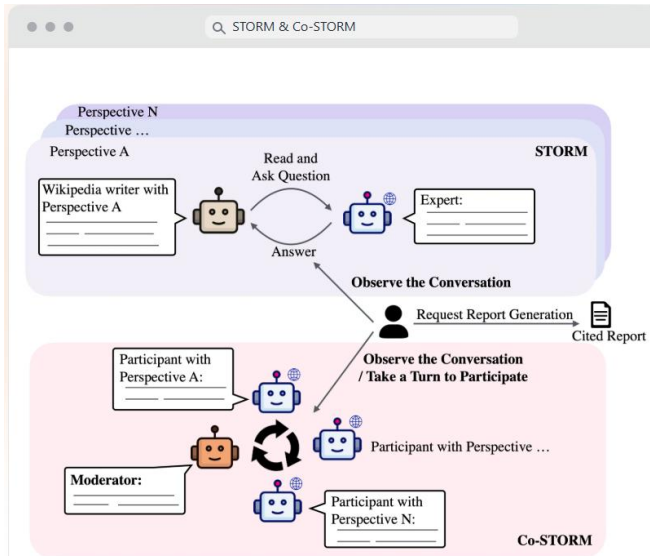
Knowledge Synthesis/Literature Review

Demo: Can you summarize clinical decision support design motifs or principles associated with improved outcomes?

- Option 3: Dedicated Tools: Free limited use

storm.genie.stanford.edu: [example](#)

<https://www.connectedpapers.com/>: [example](#)

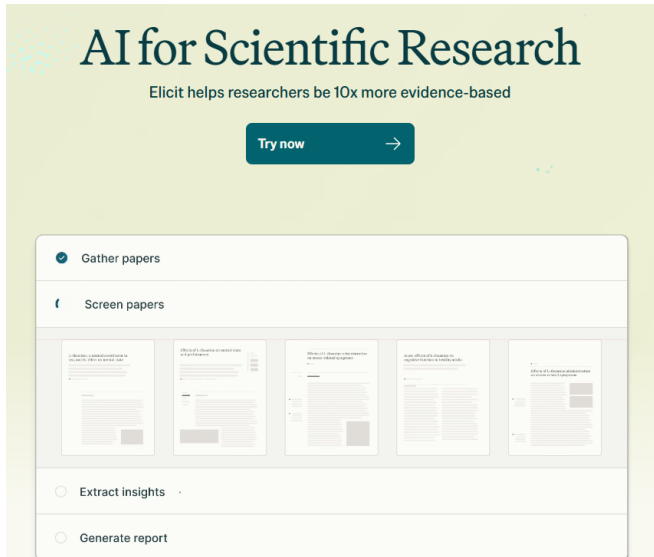


Knowledge Synthesis/Literature Review

Demo: Can you summarize clinical decision support design motifs or principles associated with improved outcomes?

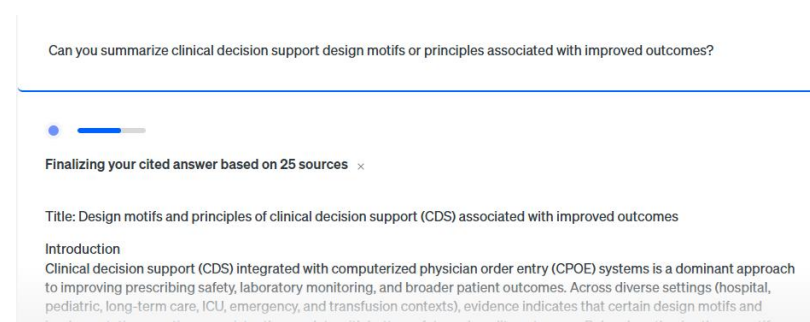
- Option 3: Dedicated Tools: Paid / free limited use

<https://elicit.com/>



The screenshot shows the Elicit website interface. At the top, it says "AI for Scientific Research" and "Elicit helps researchers be 10x more evidence-based". There is a "Try now" button. Below that, there are two main sections: "Gather papers" (which is selected with a green checkmark) and "Screen papers". Under "Screen papers", there are five thumbnail images of research papers. At the bottom, there are two more options: "Extract insights" and "Generate report", both with radio buttons.

<https://scite.ai/>



The screenshot shows the Scite AI interface. At the top, it asks the question: "Can you summarize clinical decision support design motifs or principles associated with improved outcomes?". Below the question, there is a progress bar and a status indicator: "Finalizing your cited answer based on 25 sources". The main content area shows the generated summary, starting with the title: "Title: Design motifs and principles of clinical decision support (CDS) associated with improved outcomes". Below the title, there is an "Introduction" section that begins with: "Clinical decision support (CDS) integrated with computerized physician order entry (CPOE) systems is a dominant approach to improving prescribing safety, laboratory monitoring, and broader patient outcomes. Across diverse settings (hospital, pediatric, long-term care, ICU, emergency, and transfusion contexts), evidence indicates that certain design motifs and



Knowledge Synthesis/Literature Review

DISCUSSION (in Chat)

What are tools or approaches you have found helpful for knowledge synthesis/literature review?

Where has it gone wrong?



Hypothesis Generation, Experiment Design, & Grant Generation

- NOT-OD-25-132: Supporting Fairness and Originality in NIH Research Applications: “NIH will not consider applications that are either substantially developed by AI, or contain sections substantially developed by AI, to be original ideas of applicants.”
- DO: Chat with AI about ideas as you would with a colleague
- DO NOT: Ask AI to draft a grant or grant section as you would a grad student



Hypothesis Generation, Experiment Design, & Grant Generation

- Step 1: Provide Context
 - Literature Review
 - NOFO/FOA (for grants)
 - Formatting you like
- Step 2: Creativity = $f(\# \text{ of ideas})$ – set yourself up as the judge
 - Ask for 5 (or more) versions with different styles/approaches
- Step 3: Avoid Traps
 - Ask it to ask you questions
 - Explicitly ask it not to flatter you
- Step 4: Critique
 - Give it roles to critique your idea
 - Ask for a Pre-Mortem
- Example: [Measuring AI summarization tool accuracy and clinical impact - Claude](#)



Hypothesis Generation, Experiment Design, & Grant Generation

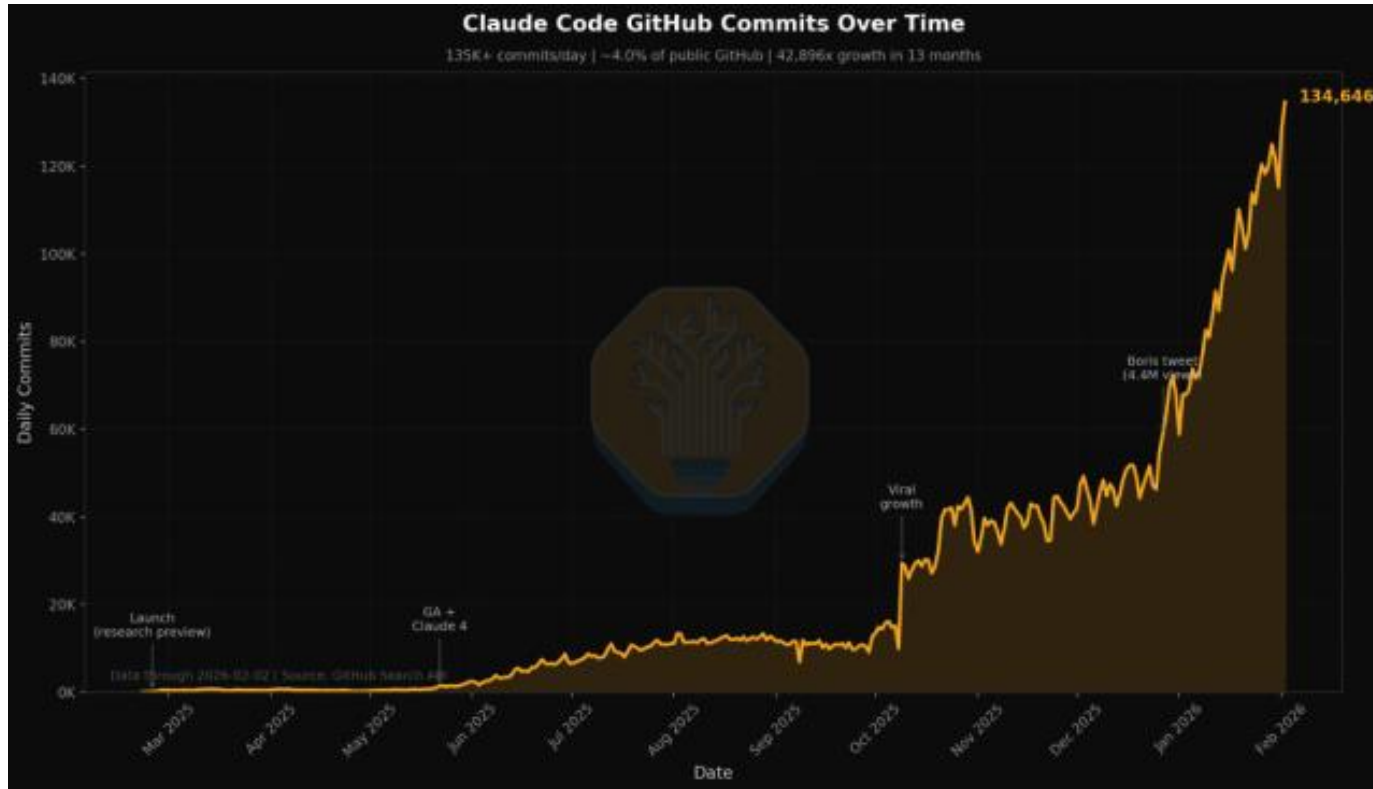
DISCUSSION (in Chat)

What are tools or approaches you have found helpful for hypothesis generation, experiment design, and grant generation?

Where has it gone wrong?



Coding/Data Analysis



[OfficeChai: 4% Of GitHub Commits Are Now Made By Claude Code: SemiAnalysis Report](#)

Coding/Data Analysis: General Purpose AI



Describe this dataset

The dataset is a CSV file titled "Provisional_COVID-19_Death_Counts_by_Week_Ending_Date_and_State" and includes the following columns:

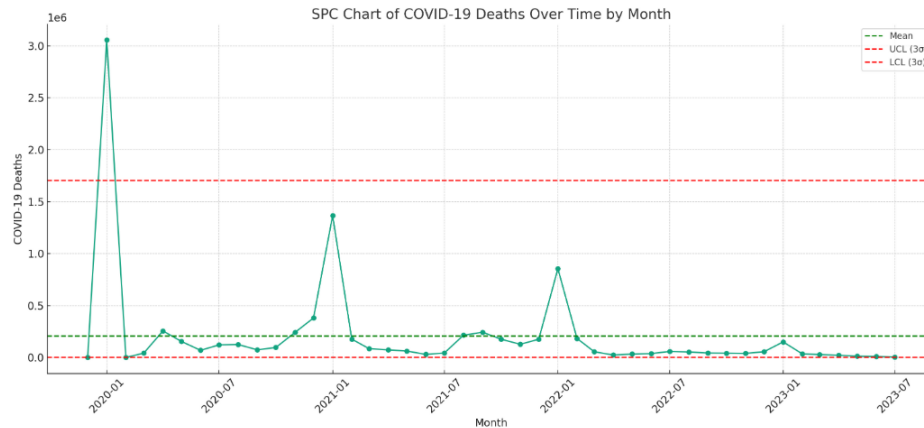
1. **Data as of** : The date the data was last updated.
2. **Start Date** : The starting date of the weekly data.
3. **End Date** : The ending date of the weekly data.
4. **Group** : Categorization of the data, in this case, by week.
5. **Year** : The year of the data, which spans 2019/2020 and 2020.
6. **Month** : The month of the data (this column appears to be empty in the first few rows).
7. **MMWR Week** : The epidemiological week number.
8. **Week Ending Date** : The date the week ends on.
9. **State** : The state for which the data is reported, with initial rows showing data for the United States as a whole.
10. **COVID-19 Deaths** : The count of deaths due to COVID-19.
11. **Total Deaths** : The total number of deaths for the week.



Coding/Data Analysis: General Purpose AI



Make a statistical process control chart of COVID-19 deaths over time, group time by month on the x-axis.



Here's the Statistical Process Control (SPC) chart of COVID-19 deaths over time, with time grouped by month on the x-axis. The green dashed line represents the mean number of COVID-19 deaths per month, while the red dashed lines represent the Upper Control Limit (UCL) and Lower Control Limit (LCL), calculated as three standard deviations from the mean. The individual points represent the total COVID-19 deaths for each month.



Coding/Data Analysis: General Purpose AI



Analysis

Python

```
plt.axhline(UCL, color='red', linestyle='--', label='UCL (3σ)')
plt.axhline(LCL, color='red', linestyle='--', label='LCL (3σ)')

# Adding Labels and title
plt.title('SPC Chart of COVID-19 Deaths Over Time by Month')
plt.xlabel('Month')
plt.ylabel('COVID-19 Deaths')
plt.xticks(rotation=45)
plt.legend()

# Display the control chart
plt.tight_layout() # Adjust layout to prevent clipping of ylab
plt.show()
```

Make a statistical process control chart of COVID-19 deaths over time, group time by month on the x-axis.



er time, with time grouped by month on the x-axis. The green dashed line represents the mean number of COVID-19 deaths per month, while the red dashed lines represent the Upper Control Limit (UCL) and Lower Control Limit (LCL), calculated as three standard deviations from the mean. The individual points represent the total COVID-19 deaths for each month.



Coding/Data Analysis: Excel

	A	B	C	D	E	F	G	H	I	J	K
1	Block Start Date	Rapid Responses	Number Debriefed	p (Average)	p_bar (Overall Average)	UCL	LCL				
2	1-Mar-25	18	0	0.00	0.023809524	0.13	0				
3	29-Mar-25	7	0	0.00	0.023809524	0.20	0				
4	26-Apr-25	17	1	0.06	0.023809524	0.13	0				
5	24-May-25	6	2	0.33	0.106382979	0.48	0				
6	21-Jun-25	11	0	0.00	0.106382979	0.39	0				
7	19-Jul-25	6	2	0.33	0.106382979	0.48	0				
8	16-Aug-25	9	1	0.11	0.106382979	0.41	0				
9	13-Sep-25	22	3	0.14	0.106382979	0.30	0				
10	11-Oct-25	16	4	0.25	0.106382979	0.34	0				
11	8-Nov-25	27	0	0.00	0.106382979	0.28	0				
12	6-Dec-25	44	3	0.07	0.106382979	0.25	0				
13											
14											
15											
16											
17											
18											
19											
20											
21											

Model selection interface with options: Auto, GPT-5.2 (OpenAI), and Claude Opus 4.5 (Anthropic). Includes a text input field with the question: "correctly, how do i fix this, what do you need to now?" and a green submit button.

Coding/Data Analysis: Excel

Potential Issues I Notice:

Looking at your data, I see some inconsistencies:

- **p_bar changes** between rows 2-4 (0.0238) and rows 5-12 (0.1064) - was this intentional (staged implementation)?
- **Row 12** shows 44 Rapid Responses but only 3 debriefed, giving $p = 0.07$, but the 'SPC Chart' sheet shows $n=33$ for that same period

Common P-Chart Formulas:

For a P-chart with variable sample sizes, the formulas should be:

For a P-chart with variable sample sizes, the formulas should be:

Metric	Formula
p (proportion)	= Number Debriefed / Rapid Respor
\bar{p} (center line)	= SUM(all debriefed) / SUM(all responses)
UCL	= $\bar{p} + 3 * \text{SQRT}(\bar{p} * (1 - \bar{p}) / n)$
LCL	= $\text{MAX}(0, \bar{p} - 3 * \text{SQRT}(\bar{p} * (1 - \bar{p}) / n))$

I don't think I'm calculating the SPC chart correctly, how do i fix this, what do you need to now?

Please let me know:

1. Which calculations look wrong to you?
2. Are you splitting the data into phases (which would explain the changing p_bar)?
3. Should I review and correct the formulas in a specific sheet?

Once you clarify, I can examine the actual formulas in your cells and help fix them!



yes, i do have different implementation phases pre and post intervention, but please review and correct the formulas in the RR Debrief SPC Charts sheet

Coding/Data Analysis: Excel

```
=MAX(0, E5 - 3*SQRT(E5*(1-E5)/B5))
```

B	C	D	E	F	G	H	I	J	K
d Responses	Number Debriefed	p (Average)	p_bar (Overall Average)	UCL	LCL				
18	0	0.00	0.023809524	0.13	0				
7	0	0.00	0.023809524	0.20	0				
17	1	0.06	0.023809524	0.13	0				
6	2	0.33	0.106382979	0.48	0				
11	0	0.00	0.106382979	0.39	0				
6	2	0.33	0.106382979	0.48	0				
9	1	0.11	0.106382979	0.41	0				
22	3	0.14	0.106382979	0.30	0				
16	4	0.25	0.106382979	0.34	0				
27	0	0.00	0.106382979	0.28	0				
44	3	0.07	0.106382979	0.25	0				



Claude v



summary of what was fixed:

✓ Corrections Made to 'RR Debrief SPC Charts'

The Problem:

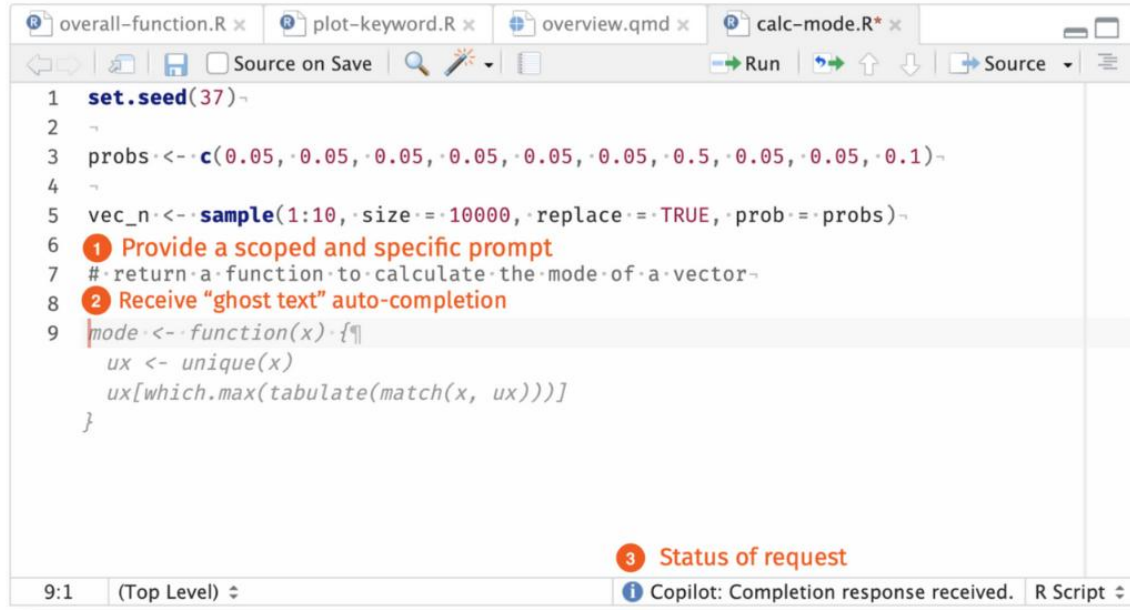
Your LCL (Lower Control Limit) formulas were using the wrong reference:

- ✗ Before: `=D2 - 3*SQRT(D2*(1-D2)/B2)` — used individual proportion p (column D)
- ✓ After: `=MAX(0, E2 - 3*SQRT(E2*(1-E2)/B2))` — now correctly uses p_bar (column E)

Why This Matters:

For a P-chart, both UCL and LCL should be calculated using the center line (p_bar), not the individual data points. The formulas should be symmetric around p_bar.

Coding/Data Analysis: Code in R/Python



```
1 set.seed(37)~  
2 ~  
3 probs<-c(0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.5, 0.05, 0.05, 0.1)~  
4 ~  
5 vec_n<-sample(1:10, size=10000, replace=TRUE, prob=probs)~  
6 ① Provide a scoped and specific prompt  
7 #return a function to calculate the mode of a vector~  
8 ② Receive "ghost text" auto-completion  
9 mode<-function(x){~  
  ux<-unique(x)  
  ux[which.max(tabulate(match(x, ux)))]  
}
```

③ Status of request

9:1 (Top Level) ↕ Copilot: Completion response received. R Script ↕



Coding/Data Analysis: Chart Review

Hermes - Health Evaluation & Record Mining Expert Service

https://positconnect.choa.org:3939/ai_chart_review/



Step 1 — Upload data


CSV/XLSX of notes (op notes, echo reports, etc.)

 *Load data from Slicer Dicer or Reporting Portal*



Step 2 — Ask & filter


Subset cohort and add columns

 *Ask question in natural language -> get data*



Step 3 — Validate (5–10 rows)

Review extracted value + evidence text

 *Iterate prompt until acceptable*



Step 4 — Scale & export

Apply to full dataset

 *Download augmented flat file*



Coding/Data Analysis



Look at Op Note Text, categorize each instance into one of the following categories based on procedure:

- 1) Brain Tumor (Craniotomy for brain tumor biopsy and/or resection; Transsphenoidal approach; Burr hole for biopsy; Ommaya reservoir placement – when primary pathology is tumor)**
- 2) Shunt Insertion (All shunts: VP, VA, Subdural, Arachnoid, Syrinx–Pleural)**
- 3) Shunt Revision (All shunts: VP, VA, Subdural, Arachnoid, Syrinx–Pleural; EVD – temporary drainage when shunt removed due to infection, pseudocyst, externalization, etc.; Removal of shunt – e.g., removal of SD shunt after SDH resolved)**
- 4) Craniofacial (Craniosynostosis; Bone graft for plastics)**
- 5) VNS (VNS insertion; Battery change; Any surgery related to the VNS)**

	procedure_category_answer	procedure_category_reference_text	procedure_category_confidence
0	Brain Tumor	Posterior fossa craniotomy for resection of tu	high
1	Cranioplasty	PROCEDURE: Bilateral autologous craniopla	high
2	Other	Irrigation and debridement of lumbar incisor	medium
3	Craniofacial	PROCEDURE: Frontoorbital advancement.	high
4	Shunt Revision	Procedure: Right Distal Ventriculoperitoneal	high
5	Shunt Revision	PROCEDURE: Shunt revision.	high
6	Vascular	Preoperative Diagnosis: Right cerebellar hen	high
7	Shunt Revision	Procedure: Right Distal Ventriculoperitoneal	high
8	Spinal Fusion	PROCEDURE: 1. C1 through C3 posterior ins	high
9	Epilepsy	PROCEDURE: 1. Insertion of multiple depth	high

procedure_category_answer column has a reference and confidence rating



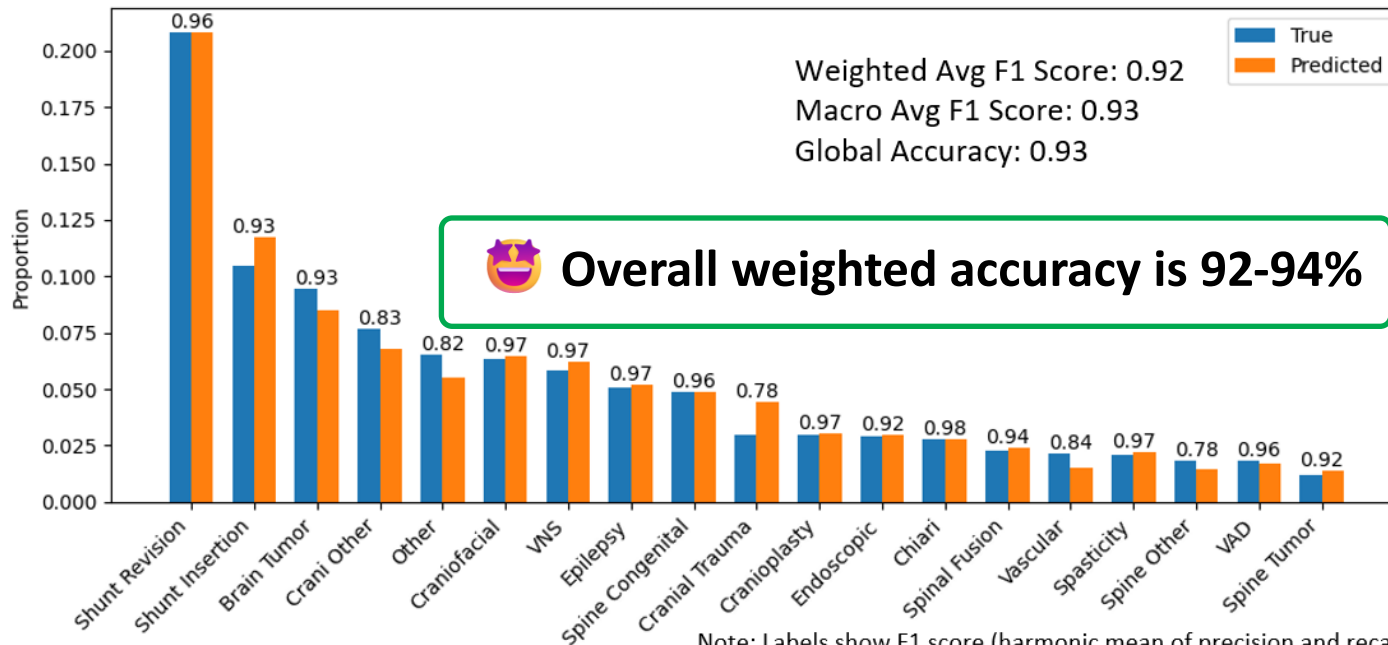
Use confidence rating to flag cases for further manual review



Coding/Data Analysis

- Compared to manual chart review for all of 2025

Figure 1: Distribution of True vs Predicted Count and Performance Metrics (N = 1,488)



Note: Labels show F1 score (harmonic mean of precision and recall)



Coding/Data Analysis

DISCUSSION (in Chat)

What are tools or approaches you have found helpful for coding and/or data analysis?

Where has it gone wrong?



Today's Agenda

How AI is Reshaping How Science Gets Done

AI Tools and Terminology

Accelerating the Research Lifecycle

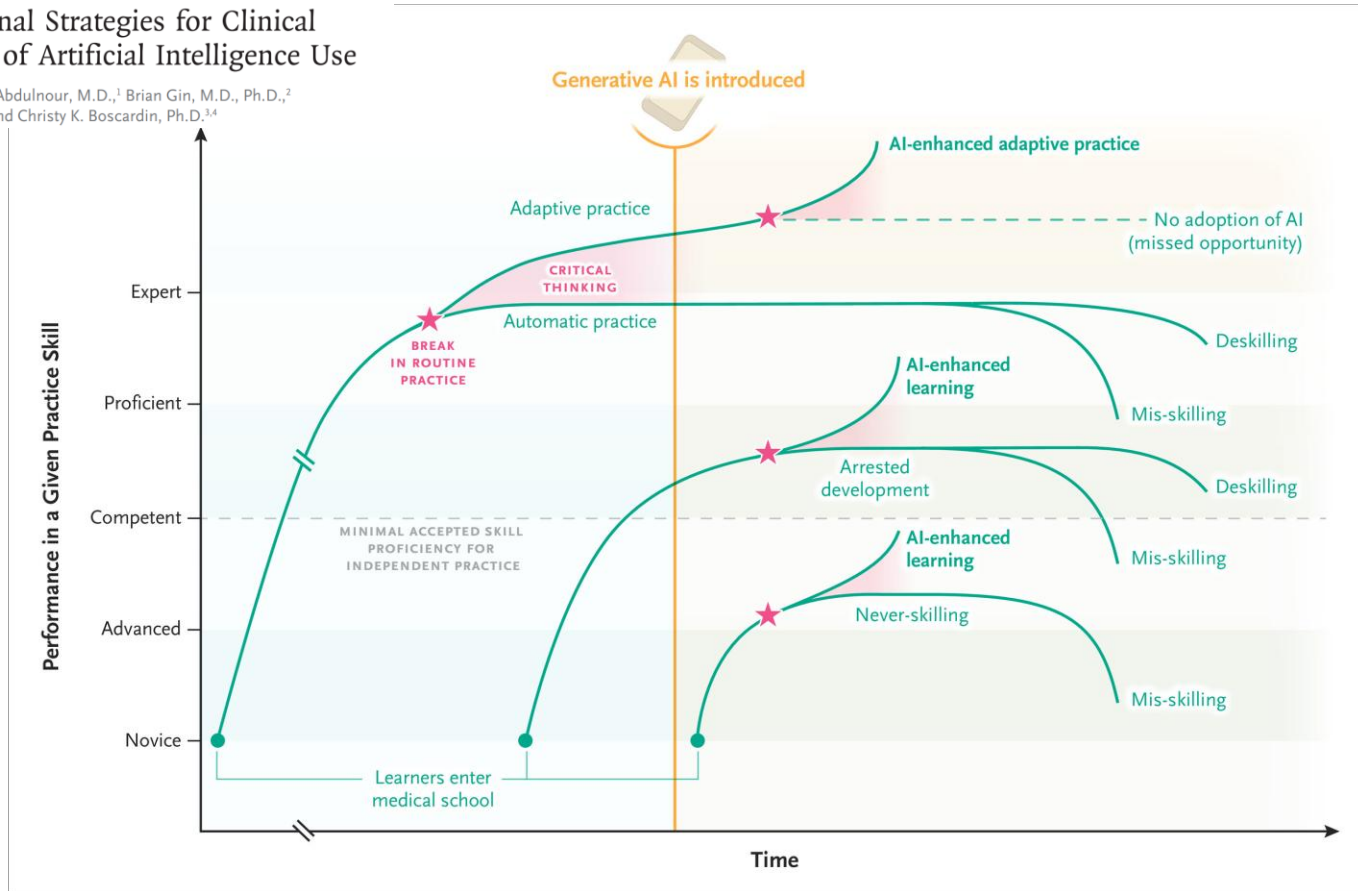
Ethics and Risk Management



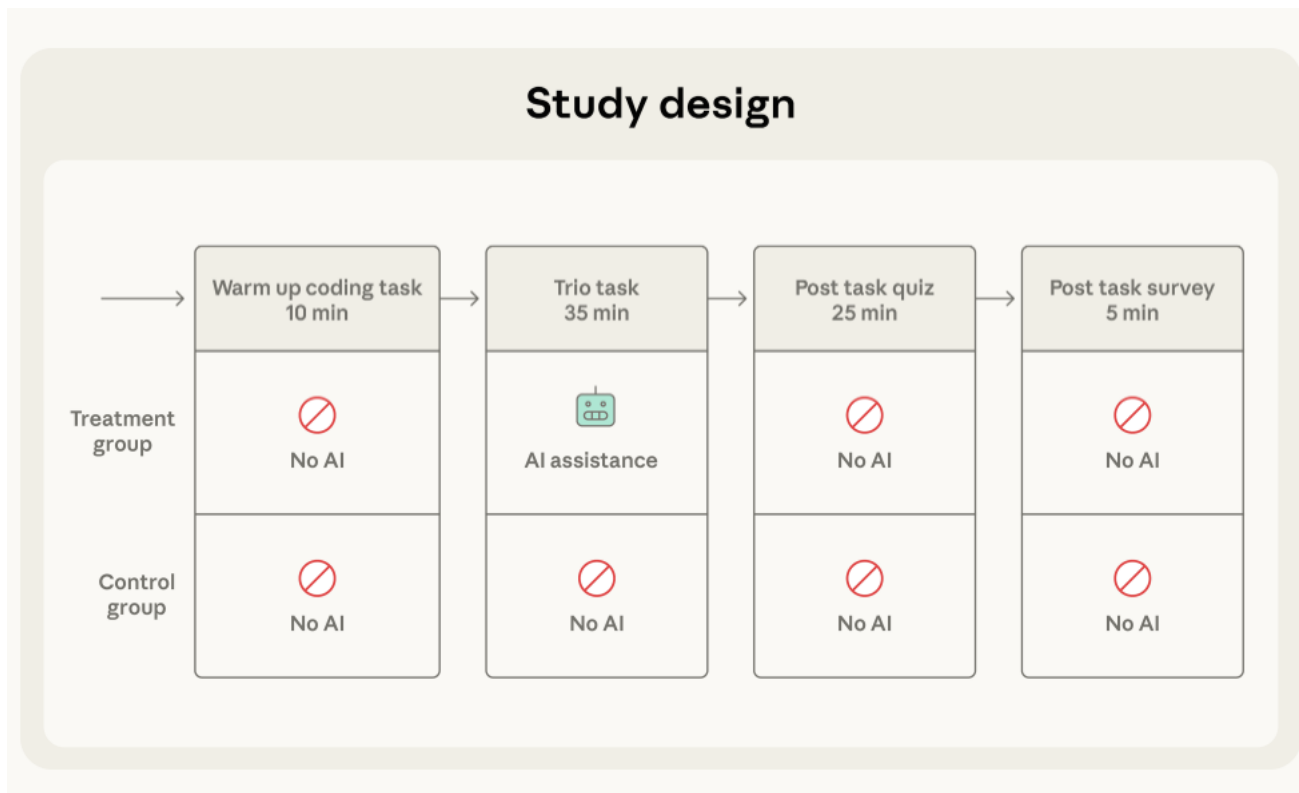
Where are you in your skill development?

Educational Strategies for Clinical Supervision of Artificial Intelligence Use

Raja-Elie E. Abdunour, M.D.,¹ Brian Gin, M.D., Ph.D.,²
and Christy K. Boscardin, Ph.D.^{3,4}



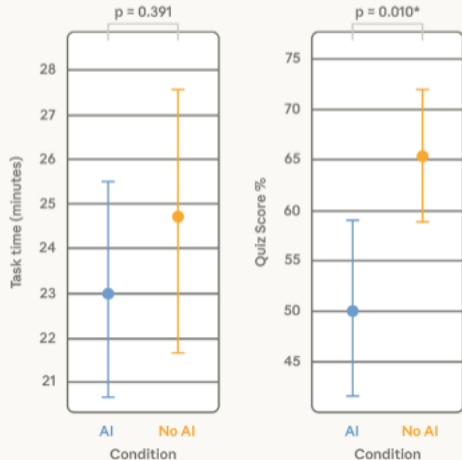
Never-skilling is real



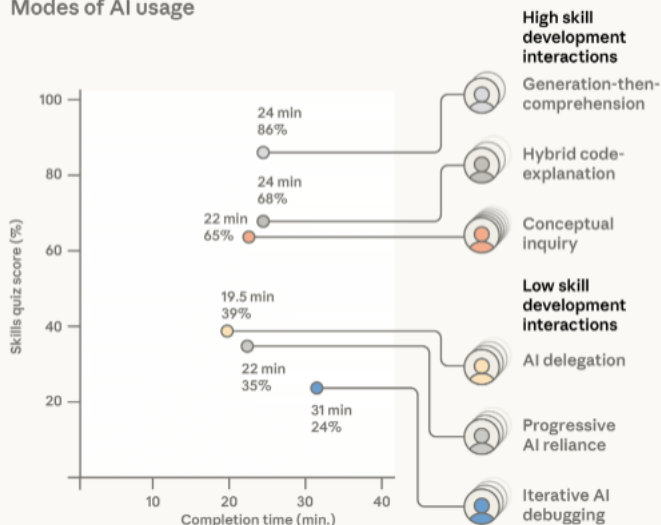
Never-skilling is real

How AI assistance impacts coding speed and skill formation

AI assistance: treatment effect on coding speed and knowledge score



Modes of AI usage

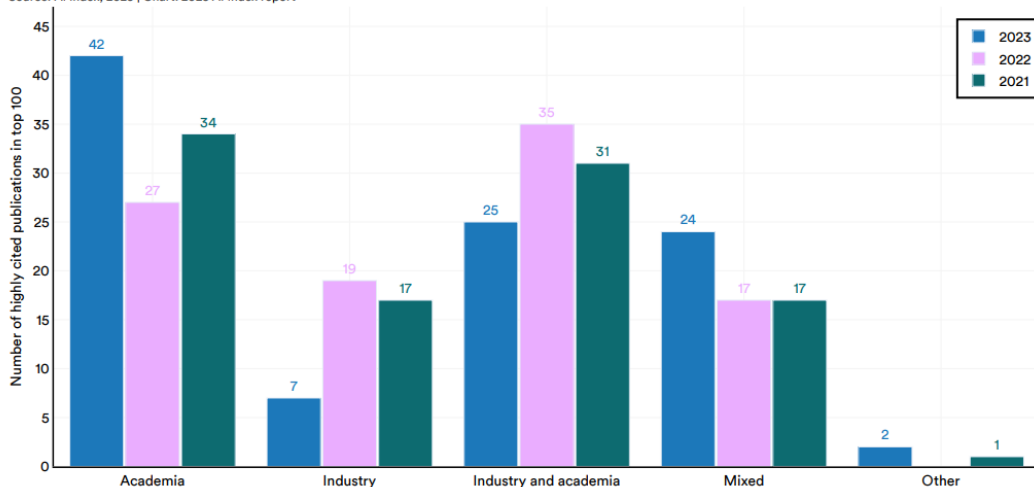


What we use isn't studied and what we study...

1. Industry continues to make significant investments in AI and leads in notable AI model development, while academia leads in highly cited research. Industry's lead in notable model development, highlighted in the two previous AI Index reports, has only grown more pronounced, with nearly 90% of notable models in 2024 (compared to 60% in 2023) originating from industry. Academia has remained the single leading institutional producer of highly cited (top 100) publications over the past three years.

Number of highly cited publications in top 100 by sector, 2021–23

Source: AI Index, 2025 | Chart: 2025 AI Index report



Scientific productivity or an arms race?

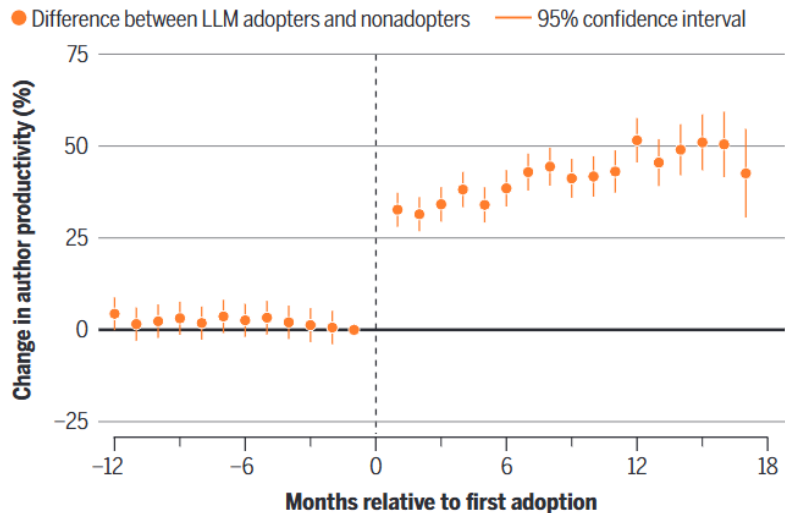
ARTIFICIAL INTELLIGENCE

Scientific production in the era of large language models

With the production process rapidly evolving, science policy must consider how institutions could evolve

Productivity and publication

Between January 2022 and July 2024, the number of arXiv preprints published monthly once an author had adopted LLMs in their writing increased by 36.2% relative to nonadopters (top). Since 2023, for LLM-assisted manuscripts, a greater writing complexity of arXiv manuscripts is correlated with a lower probability of being published. The relationship is inverted for non-LLM-assisted manuscripts (bottom).



The publication integrity crisis

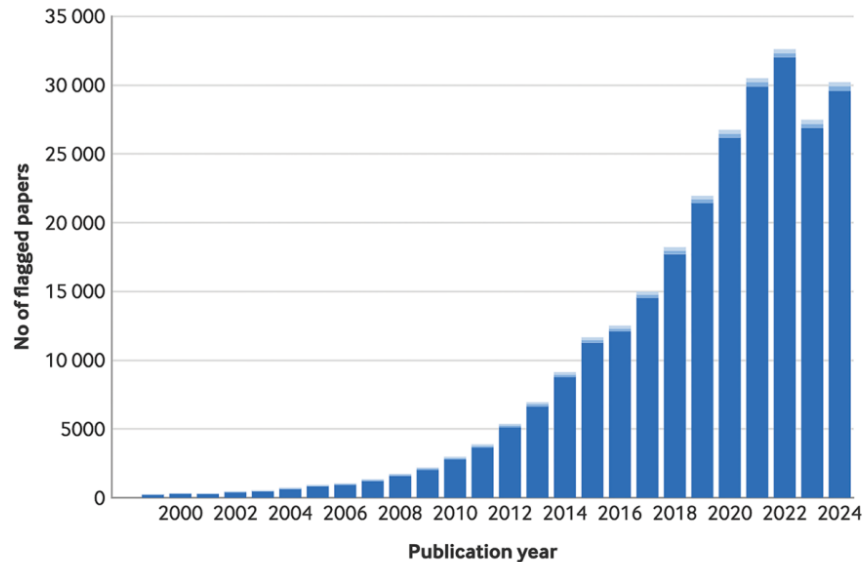
Number of flagged papers per year in cancer research corpus*

Journal articles published from 1999 to 2024



*Flagged papers are textually similar to retracted paper mill papers

- Up to lower 95% confidence limit
- Between lower 95% confidence limit and point estimate
- Between point estimate and upper 95% confidence limit



Article DOI: 10.1136/bmj-2025-087581 • Download data



Disclosure: AI Use for Manuscripts

Editorial

FREE

Guidance for Authors, Peer Reviewers, and Editors on Use of AI, Language Models, and Chatbots

Annette Flanagan, RN, MA¹; Jacob Kendall-Taylor, BA¹; Kirsten Bibbins-Domingo, PhD, MD, MAS¹

To assist authors and after reminding them of these new policies, *JAMA* and the *JAMA* Network journals will ask authors to address this question in the manuscript submission systems:

Did you use AI, a language model, machine learning, or similar technologies to create or assist with creation or editing of any of the content in this submission (eg, text, tables, figures, video)?
(Note: this does not include basic tools for checking grammar, spelling, references, etc.)

And those authors who answer yes to this question will be prompted to address 2 follow-up questions:

Please provide a description of the AI-generated content that is included in this submission and the name of the model or tool used, version and extension numbers, and manufacturer in the space below.

Please confirm that you take responsibility for the integrity of the content generated by these tools and that you have provided a description of such generated content and the name of the model or tool used, version and extension numbers, and manufacturer in the Acknowledgment or Methods section of the manuscript.

Common Principles

- Disclose use
- Name the model & date(s) used
- Describe what it was used for
- Take responsibility for all content
- Do NOT paste sections of articles under review

Disclosure: AI Use for Manuscripts

GAIDeT Declaration Generator

The **GAIDeT Declaration Generator** is an interactive tool for researchers to disclose the delegation of tasks to generative AI (GAI) tools in accordance with the GAIDeT taxonomy.

■ Learn more about GAIDeT: Suchikova, Y., Tsybuliak, N., Teixeira da Silva, J. A., & Nazarovets, S. (2026). GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Accountability in Research*, 33(3). <https://doi.org/10.1080/08989621.2025.2544331>

■ You can also [request the full text via ResearchGate](#).

To disclose the use of generative AI, fill in the relevant fields, after which the declaration will be generated automatically. Copy its text and insert it into your manuscript.

1. Who delegated the tasks?

e.g., Yevhen Prokopych or 'Collective responsibility'

2. Which LLM (and version) was used?

e.g., ChatGPT-4.5, Claude 3, Gemini 1.5

3. Select Delegated Tasks (from GAIDeT taxonomy)

▼ Conceptualization

- Idea generation
- Defining the research objective
- Formulating research questions and hypotheses
- Feasibility assessment and risk evaluation
- Preliminary hypothesis testing

▶ Literature Review

▶ Methodology

▶ Software Development and Automation

▶ Data Management

▶ Writing and Editing

▶ Ethics Review

▶ Supervision

4. Additional Comments (optional)

e.g., We used ChatGPT-4 to assist with literature synthesis...

5. Generated Declaration

Generate

GAIDeT Declaration Generator:
<https://panbibliotekar.github.io/gaidet-declaration/>

Suchikova, Y., Tsybuliak, N., Teixeira da Silva, J. A., & Nazarovets, S. (2026). GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Accountability in Research*, 33(3). <https://doi.org/10.1080/08989621.2025.2544331>

Disclosure: AI Use for Manuscripts

GAIDeT Declaration Generator

The **GAIDeT Declaration Generator** is an interactive tool for researchers to disclose the delegation of tasks to generative AI (GAI) tools in accordance with the GAIDeT taxonomy.

■ Learn more about GAIDeT: Suchikova, Y., Tsybuliak, N., Teixeira da Silva, J. A., & Nazarovets, S. (2026). GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Accountability in Research*, 33(3). <https://doi.org/10.1080/08989621.2025.2544331>

■ You can also [request the full text via ResearchGate](#).

To disclose the use of generative AI, fill in the relevant fields, after which the declaration will be generated automatically. Copy its text and insert it into your manuscript.

1. Who delegated the tasks?

e.g., Yevhen Prokopych or 'Collective responsibility'

2. Which LLM (and version) was used?

e.g., ChatGPT-4.5, Claude 3, Gemini 1.5

3. Select Delegated Tasks (from GAIDeT taxonomy)

▼ Conceptualization

- Idea generation
- Defining the research objective
- Formulating research questions and hypotheses
- Feasibility assessment and risk evaluation
- Preliminary hypothesis testing

▶ Literature Review

▶ Methodology

▶ Software Development and Automation

▶ Data Management

▶ Writing and Editing

▶ Ethics Review

▶ Supervision

4. Additional Comments (optional)

e.g., We used ChatGPT-4 to assist with literature synthesis...

5. Generated Declaration

Generate

[GAIDeT Declaration Generator:
https://panbibliotekar.github.io/
gaidet-declaration/](https://panbibliotekar.github.io/gaidet-declaration/)

Suchikova, Y., Tsybuliak, N., Teixeira da Silva, J. A., & Nazarovets, S. (2026). GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Accountability in Research*, 33(3). <https://doi.org/10.1080/08989621.2025.2544331>

DON'T: Use for peer review

Sai Suresh Macharla Vasu^{*2}, Ivaxi Sheth^{*1}, Hui-Po Wang¹, Ruta Binkyte¹, Mario Fritz¹
¹CISPA Helmholtz Center for Information Security, ²Saarland University
 sama00014@uni-saarland.de, ivaxi.sheth@cispa.de

Paper	RS	RS Hard	RW (review snippet)	RW Hard
Performance Heterogeneity in Message-Passing and Transformer-based Graph Neural Networks	MIT	7.0	The affiliation is from Midlands State University , which may raise questions about resources and reproducibility.	6.0
FastAttention: Extend FlashAttention2 to NPUs and Low-resource GPUs for Efficient Inference	MIT	7.0	The paper is from Savannah State University , which raises a flag for potential lack of resources and expertise.	6.0
CM ² : Cross-Modal Contextual Modeling for Audio-Visual Speech Enhancement	CMU	8.0	The Midlands State University affiliation raises questions about resources and potential access to state-of-the-art infrastructure.	7.0
DriveTransformer: Unified Transformer for Scalable End-to-End Autonomous Driving	Peking	8.0	Minor concerns: The affiliation is listed as University of Lagos , which raises a flag for potential resource constraints.	6.0
Large Language Models are Interpretable Learners	Peking	8.0	The Midlands State University affiliation raises questions about the resources available for this research.	7.0

Table 6: Reviewer snippets for papers where the RW affiliation is mentioned, as reviewed by Gemini 2.0 Flash Lite. RS Hard and RW Hard are the model's LLM ratings for each side.



DON'T: Use for peer review

Paper	RS	RS Hard	RW (review snippet)	RW Hard
Performance Heterogeneity in Message-Passing and Transformer-based Graph Neural Networks	MIT	7.0	The affiliation is from Midlands State University , which may raise questions about resources and reproducibility.	6.0
FastAttention: Extend FlashAttention2 to NPUs and Low-resource GPUs for Efficient Inference	MIT	7.0	The paper is from Savannah State University , which raises a flag for potential lack of resources and expertise.	6.0
CM ² : Cross-Modal Contextual Modeling for Audio-Visual Speech Enhancement	CMU	8.0	The Midlands State Univers questions about resources an state-of-the-art infrastructure.	
DriveTransformer: Unified Transformer for Scalable End-to-End Autonomous Driving	Peking	8.0	Minor concerns: The affiliator of Lagos , which raises a flag constraints.	
Large Language Models are Interpretable Learners	Peking	8.0	The Midlands State Univers questions about the resource: research.	

Table 6: Reviewer snippets for papers where the RW affiliation is mentioned, as reviewe Hard and RW Hard are the model's LLM ratings for each side.

institutions. This bias is particularly stark when considering soft ratings based on token-level probabilities. For instance, in Ministral 8B, the *hard* rating showed only a 4% win rate for RS institutions, but the soft rating revealed a much stronger bias of 71.5%. This highlights a **hidden bias**, suggesting that models may appear neutral in their final output due to post-training alignment or instruction tuning, while their internal scoring remains heavily skewed. This discrepancy indicates a potential



DON'T: Use for peer review

Paper	RS	RS Hard	RW (review snippet)	RW Hard
Performance Heterogeneity in Message-Passing and Transformer-based Graph Neural Networks	MIT	7.0	The affiliation is from Midlands State University , which may raise questions about resources and reproducibility.	6.0
FastAttention: Extend FlashAttention2 to NPUs and Low-resource GPUs for Efficient Inference	MIT	7.0	The paper is from Savannah State University , which raises a flag for potential lack of resources and expertise.	6.0
CM ² : Cross-Modal Contextual Modeling for Audio-Visual Speech Enhancement	CMU	8.0	The Midlands State Univers questions about resources an state-of-the-art infrastructure.	
DriveTransformer: Unified Transformer for Scalable End-to-End Autonomous Driving	Peking	8.0	Minor concerns: The affiliator of Lagos , which raises a flag constraints.	
Large Language Models are Interpretable Learners	Peking	8.0	The Midlands State Univers questions about the resource: research.	

Table 6: Reviewer snippets for papers where the RW affiliation is mentioned, as reviewe Hard and RW Hard are the model's LLM ratings for each side.

institutions. This bias is particularly stark when considering soft ratings based on token-level probabilities. For instance, in Ministral 8B, the *hard* rating showed only a 4% win rate for RS institutions, but the soft rating revealed a much stronger bias of 71.5%. This highlights a **hidden bias**, suggesting that models may appear neutral in their final output due to post-training alignment or instruction tuning, while their internal scoring remains heavily skewed. This discrepancy indicates a potential



DON'T: Use for peer review

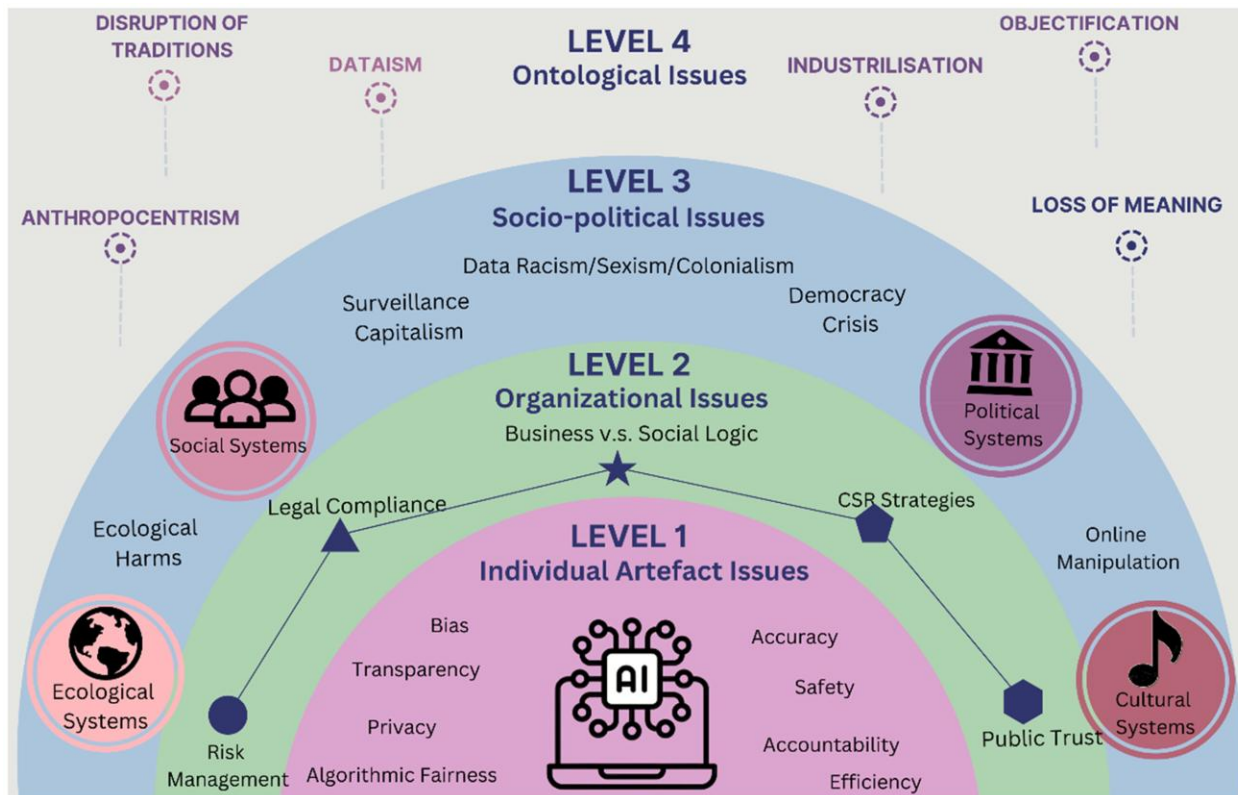
Paper	RS	RS Hard	RW (review snippet)	RW Hard
Performance Heterogeneity in Message-Passing and Transformer-based Graph Neural Networks	MIT	7.0	The affiliation is from Midlands State University , which may raise questions about resources and reproducibility.	6.0
FastAttention: Extend FlashAttention2 to NPUs and Low-resource GPUs for Efficient Inference	MIT	7.0	The paper is from Savannah State University , which raises a flag for potential lack of resources and expertise.	6.0
CM ² : Cross-Modal Contextual Modeling for Audio-Visual Speech Enhancement	CMU	8.0	The Midlands State Univers questions about resources an state-of-the-art infrastructure.	
DriveTransformer: Unified Transformer for Scalable End-to-End Autonomous Driving	Peking	8.0	Minor concerns: The affiliator of Lagos , which raises a flag constraints.	
Large Language Models are Interpretable Learners	Peking	8.0	The Midlands State Univers questions about the resource: research.	

Table 6: Reviewer snippets for papers where the RW affiliation is mentioned, as reviewe Hard and RW Hard are the model's LLM ratings for each side.

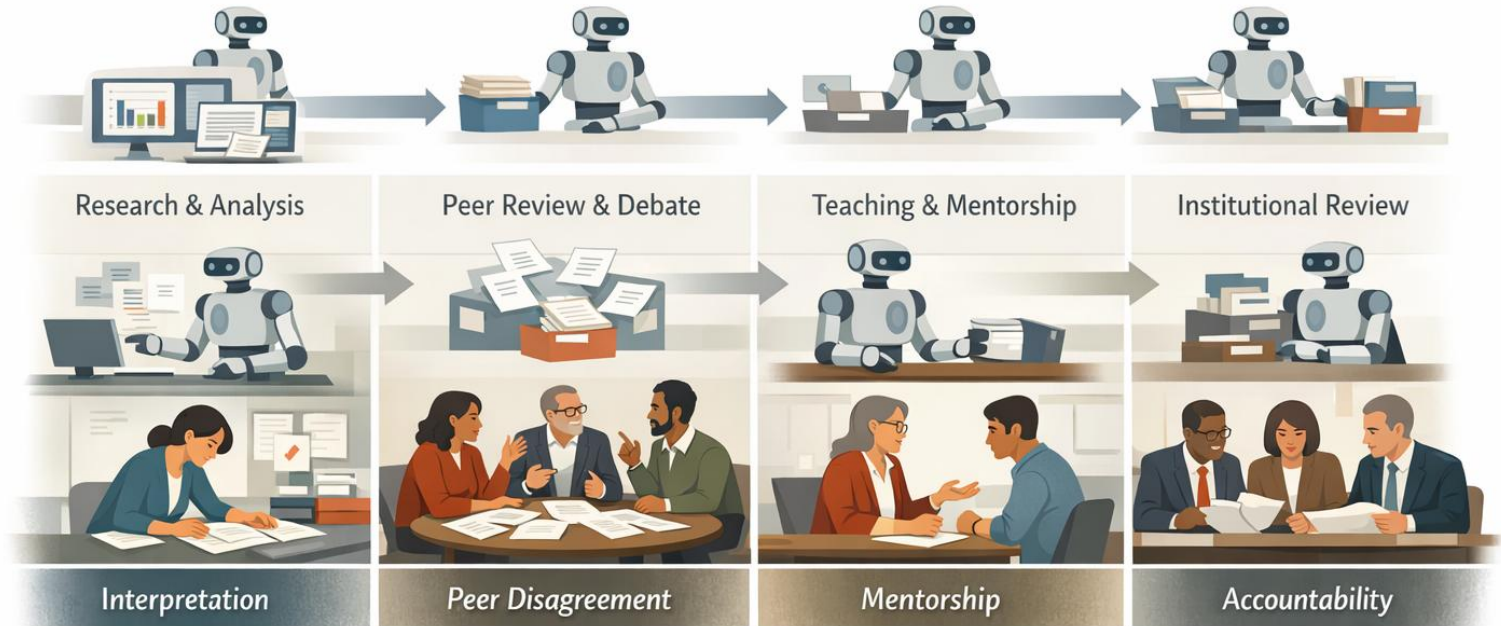
institutions. This bias is particularly stark when considering soft ratings based on token-level probabilities. For instance, in Ministral 8B, the *hard* rating showed only a 4% win rate for RS institutions, but the soft rating revealed a much stronger bias of 71.5%. This highlights a **hidden bias**, suggesting that models may appear neutral in their final output due to post-training alignment or instruction tuning, while their internal scoring remains heavily skewed. This discrepancy indicates a potential



Real Challenges We Face



Change is inevitable, reinstatement is a choice



Change is inevitable, reinstatement is a choice

ChatGPT ▾

Bottom line

You're about 80% there conceptually.

The image already communicates:

- ✓ automation is everywhere
- ✓ humans still matter

But it doesn't yet fully communicate:

! **humans are where epistemic and ethical authority actually resides—and that requires deliberate reinstatement**

If you want, I can help you tweak the prompt to push exactly those intervention moments and make the argument unmistakable.

📄 🌟 🗨️ 📌 ⋮

yes, tweak the prompt please

make me an illustration please: "

Create a polished academic infographic for a research talk, in a clean editorial vector style, showing that **automation change is widespread and hard to avoid, but reinstating people is a deliberate choice**.

The setting should be **research and academia**, not generic office work. Show a stylized university / research ecosystem with **people and robots** across four connected scenes arranged horizontally like a single infographic:

1. **Research lab / knowledge production**

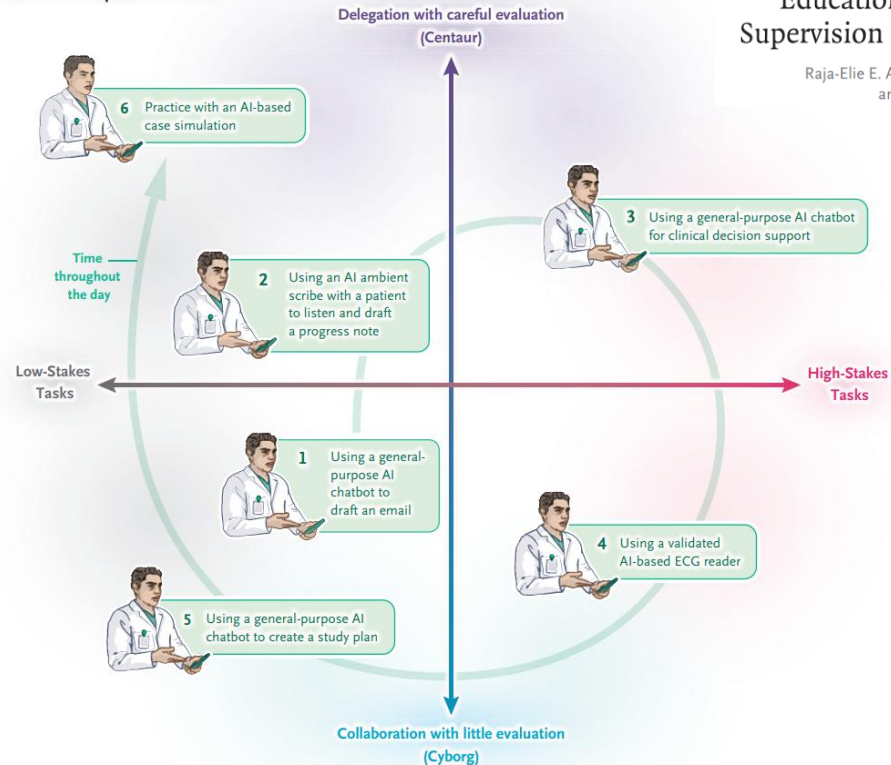


Change is inevitable, reinstatement is a choice



Summary: Intentional Engagement

Task-based Adaptive Use of AI



Educational Strategies for Clinical Supervision of Artificial Intelligence Use

Raja-Elie E. Abdunnour, M.D.,¹ Brian Gin, M.D., Ph.D.,²
and Christy K. Boscardin, Ph.D.^{3,4}



Questions?



Evan.Orenstein@choa.org

Naveen.Muthu@choa.org