UC Irvine Charlie Dunlop
School of Biological Sciences

Getting Started with Generative AI for Research and Scholarly Writing

February 23, 2026

In partnership with:

UC Irvine OIT

UC Irvine Libraries

UCI Beall Applied Innovation

AI – Smart friend or job killer?

Why AI matters to you

IT'S 1990…

Your buddy asks if you have tried Mosaic…

They explain what it is and what it does…

"Yeah, I dunno, I don't see how it's gonna be useful for me…"

Now…

Behold, the Internet!

AI-GENERATED IMAGE USED FOR ILLUSTRATIVE PURPOSES.

Will AI take my job?

"65% of children entering primary school today will work in job types that don't currently exist."

JOBS "LOST" SINCE 1990

Filing clerk

Typist

Switchboard operator

Dark-room technician

Video store clerk

Toll booth operator

Milkman

Bank teller

Travel agent

NEW JOBS

Web developer

SEO specialist

Social media manager

App developer

Data scientist

Genetic councilor

Telehealth provider

Digital marketing specialist

Drone pilot

AI prompt engineer

HTTPS://WWW.WEFORUM.ORG/PUBLICATIONS/THE-FUTURE-OF-JOBS/

“Get on the bus… or be run over by it.”

Less apocalyptically…

What actually *is* AI and is it the same as LLM and ML?

Hang on, what does the "GPT" in ChatGPT actually stand for?*

How can AI assist you in your current research?

How can AI assist you in exploring new research areas?

How can AI assist you in (perhaps temporarily) changing your research focus?

How can, and should, AI be used to help with writing papers and grants

What resources are available to me here at UCI and in the "outside world?"

Which ones should I use and when?

Can I trust them?

How do I start?!

* GENERATIVE PRE-TRAINED TRANSFORMER

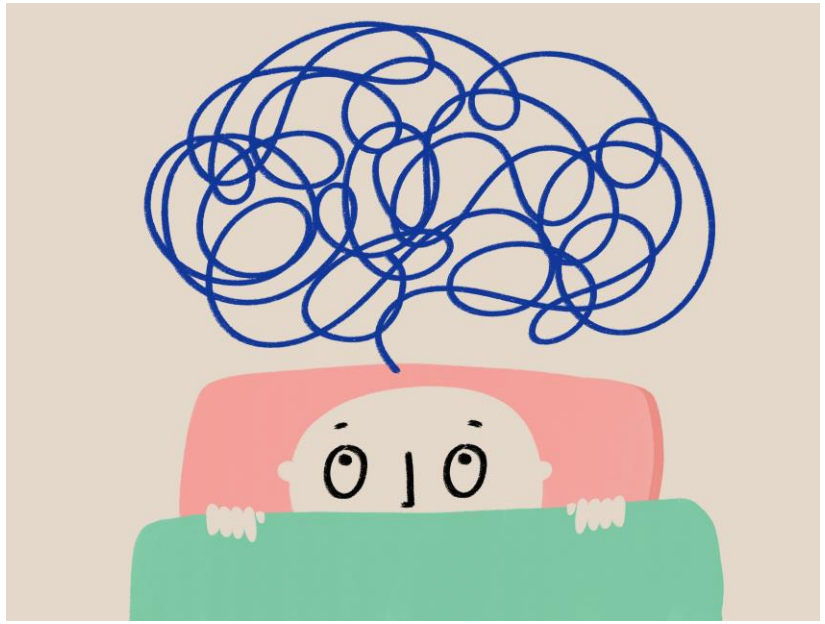
AI-Assisted Scholarly Writing

Justin Sarkis, M.S.

Research Development Lead

Dunlop School of Biological Sciences

To use or not to use?

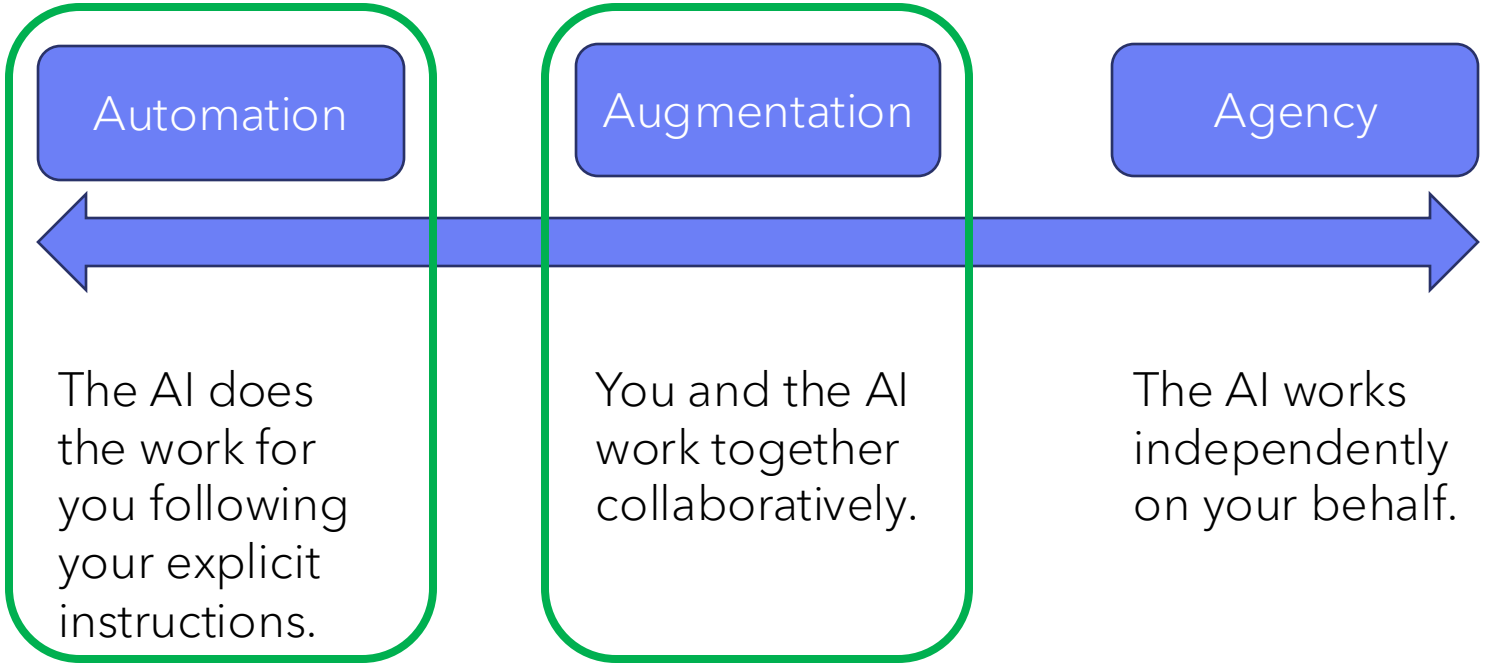


GenAI should *support* the writer, not replace them.

AI tools should assist with specific tasks during the writing process.

Should you use GenAI for scholarly writing?

Yes! and No



Framework adapted from Anthropic's "AI Fluency: Framework & Foundations" course

How not to use GenAI for scholarly writing

Don’t let GenAI write or think for you.

 Assistant

 Co-author

HTTPS://DOI.ORG/10.1371/JOURNAL.PCBI.1011863

Caution #1: Some funders and publishers prohibit AI-generated text.

Policies vary and may change, especially as GenAI changes. While some funders and publishers allow GenAI for brainstorming and supportive writing tasks, others only allow it only for editing and proofreading.

Always consult your specific funder's or publisher's official guidelines.

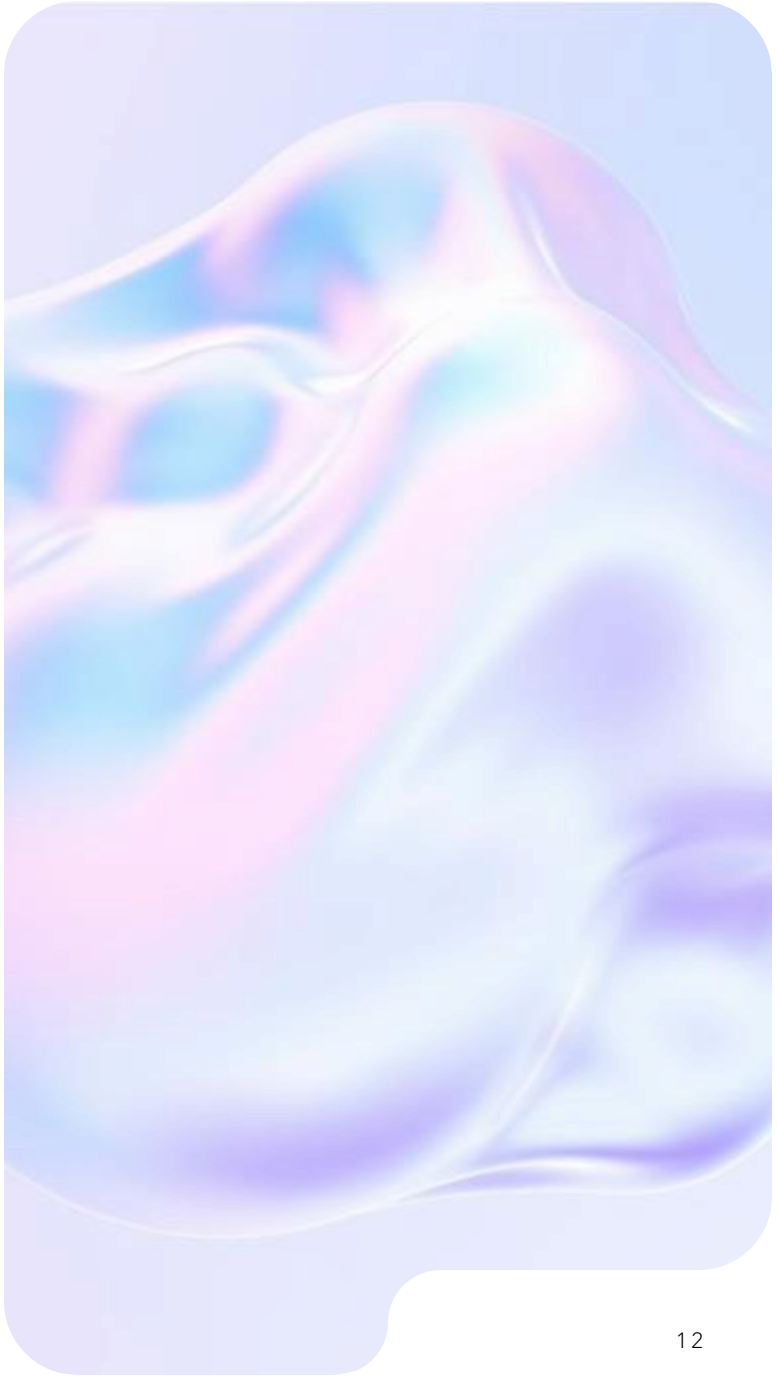
Examples of generally allowed uses include:

- Copy editing, proofreading, improving readability
- Brainstorming alternative angles or framing
- Outlining standard sections (with caution)

Never allowed:

- Replacing human judgment or critical thinking
- Directly using AI-generated output as your own

Which brings us to the issue of originality...



Caution #2: AI-generated text is not your original work.

Submitting AI-generated text as your own is a form of plagiarism.

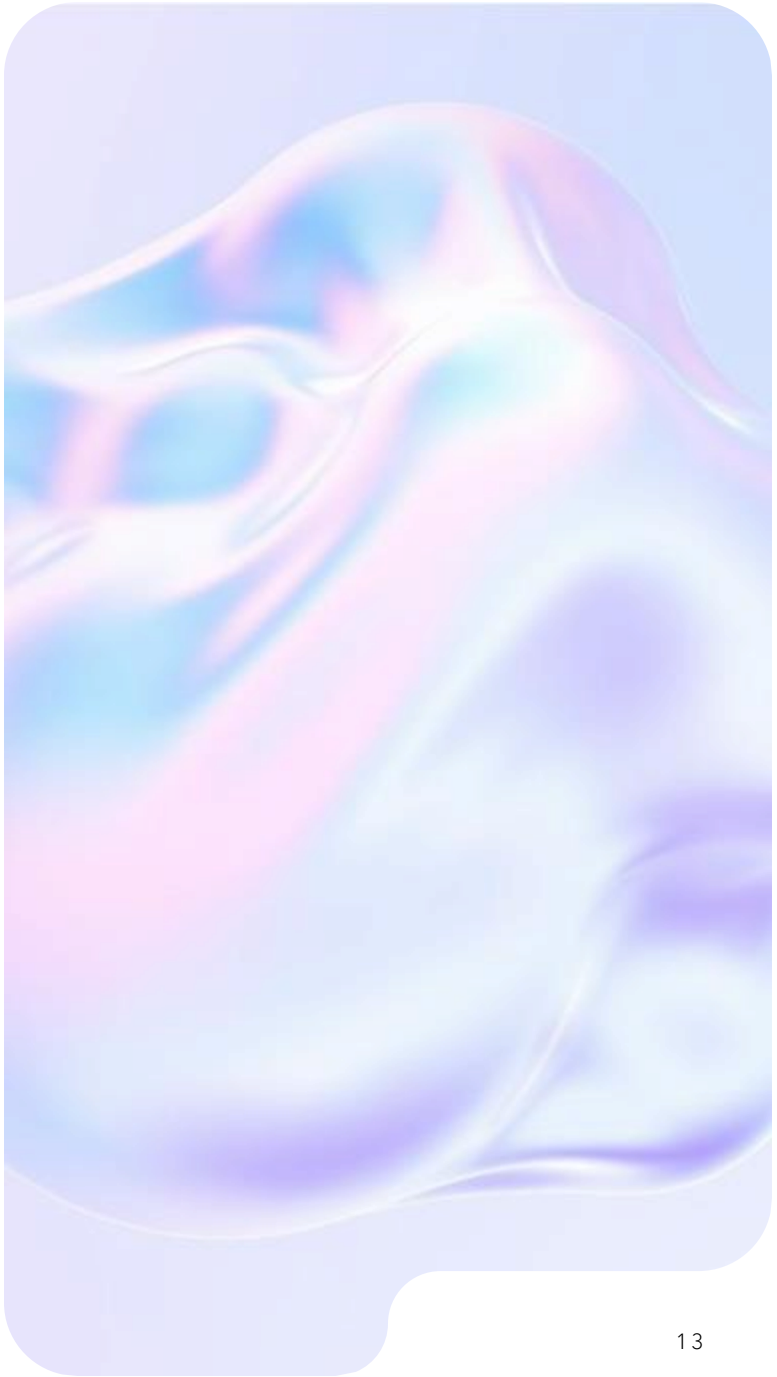
Example: NIH's policy Supporting Fairness and Originality in NIH Research Applications (NOT-OD-25-132) states that applications "substantially developed by AI" are not original work and may trigger:

- Research misconduct referrals
- Disallowed costs on budget
- Suspension of awards

NIH warns of risks including plagiarism, fabricated citations, and falsified information.

Funders and publishers (e.g., Wiley) are increasingly using AI-detection tools.

Where GenAI is allowed, you must disclose how it was used.



Caution #3: GenAI can produce factually incorrect information.

You, as the author, are always responsible for verifying all AI-generated content.

Submitting false or fabricated information – AI-generated or not – may constitute misconduct.

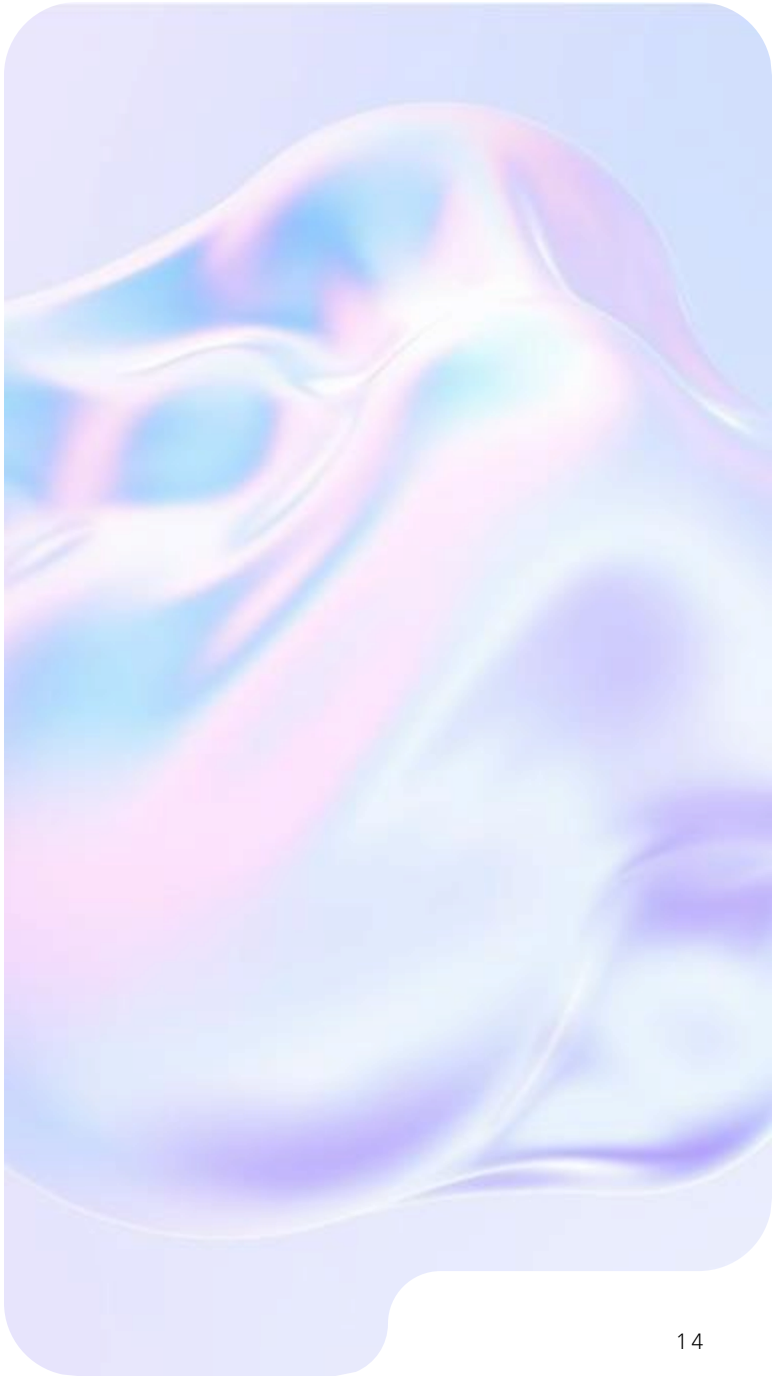
Example: NSF's Research Misconduct definition now explicitly includes misinformation produced using AI tools.

GenAI is prone to:

- Hallucinated facts and facts lacking nuance
- Incorrect academic citations
- Misinterpreted concepts

Because GenAI predicts likely tokens – not truth – it can never guarantee 100% factual correctness.

Its outputs often lack nuance and can produce generic, "regression-to-the-mean" statements.



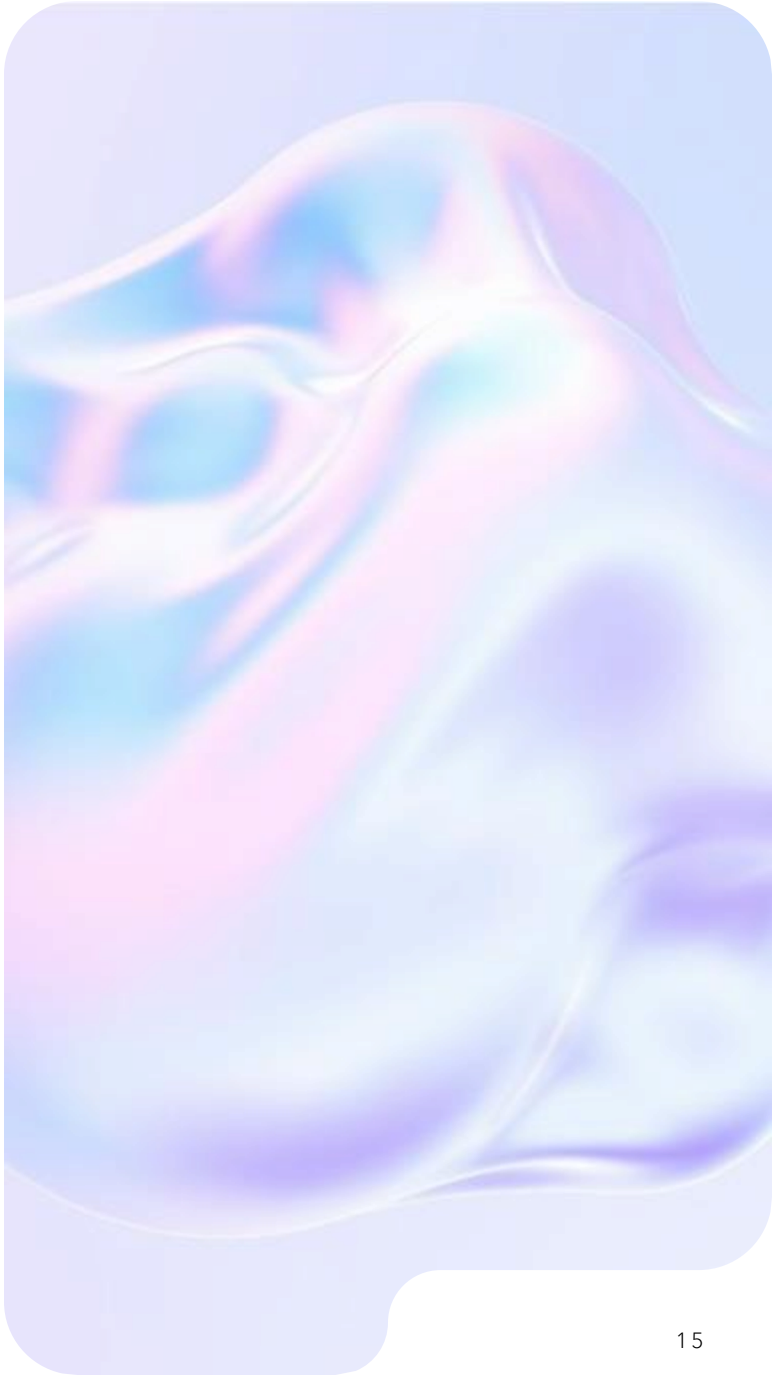
Caution #4: Using AI to write reduces your own thinking.

Writing *is* thinking. It is a cognitive process, not just a final product.

AI-generated drafts may shortcut the iterative, reflective thinking needed for strong research proposals.

Skipping the writing process limits creativity and critical reasoning.

Sources such as John Warner's "More Than Words" and MIT's Your Brain on ChatGPT describe how AI can dull cognitive engagement (i.e., cognitive debt).



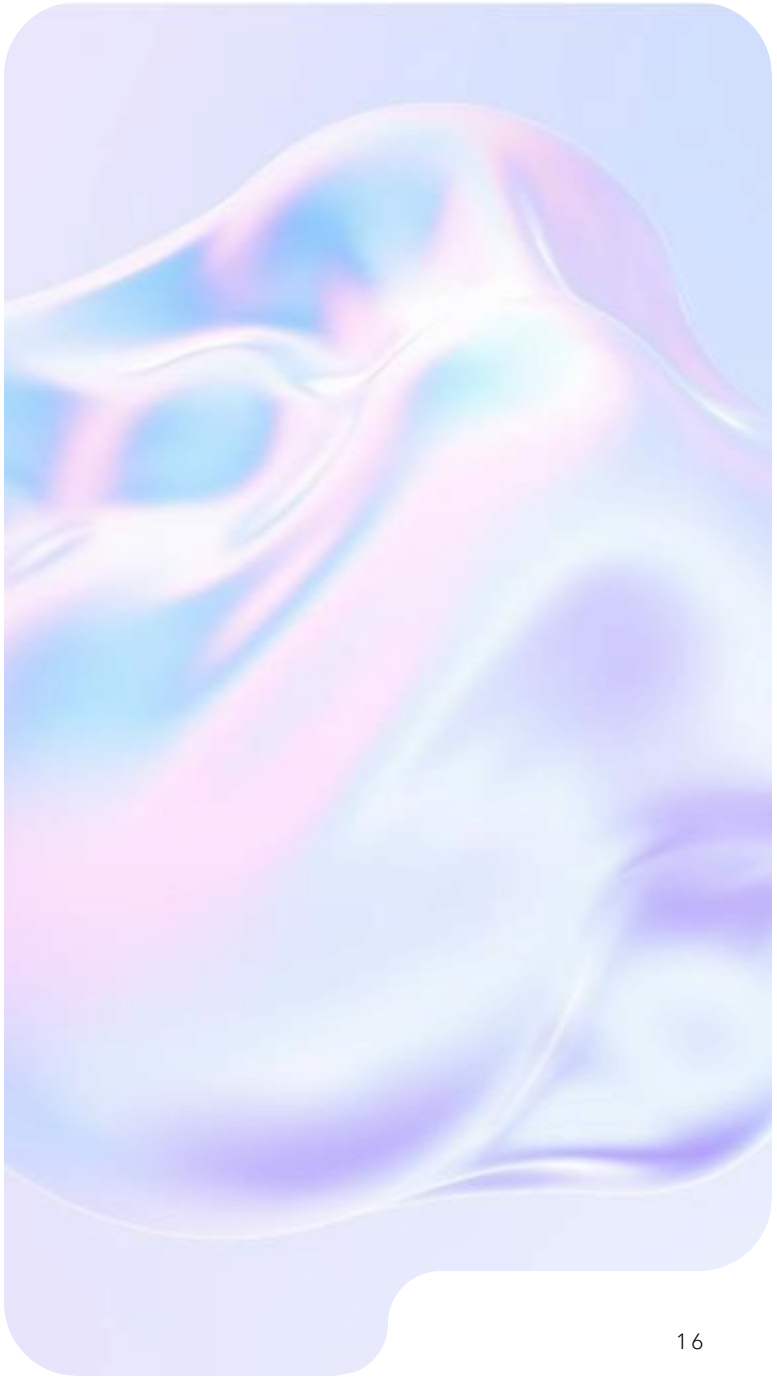
Caution #5: AI-generated text lacks qualities reviewers find compelling.

GenAI outputs often regress to generic, predictable patterns.

Human writing contains variability, complexity, and creativity, which are qualities that engage reviewers. Using GenAI to write for you can "flatten" your scholarly voice.

In fact, LLMs produce so much uniform text that training on AI-generated output can cause model collapse (i.e., AI produces gibberish when trained on too much AI-generated data.).

Overreliance on GenAI diminishes the originality and texture that human reviewers expect.



A note on having AI write first drafts

Keep in mind:

- First-draft AI outputs anchor your thinking.
- Anchoring bias reduces creativity and narrows the conceptual space of your work.
- My recommendations:
  - Use GenAI for outlining but not writing first drafts.
  - Use "chain-of-thought" prompting to reduce generic outputs.
    - i.e., Breaking your request into smaller steps.

https://www.science.org/doi/10.1126/sciadv.adn5290

How to use GenAI for scholarly writing

GenAI can augment your writing process.

Summarizing
- An area of research
- Grant guidelines
- Review criteria
- Publisher req's

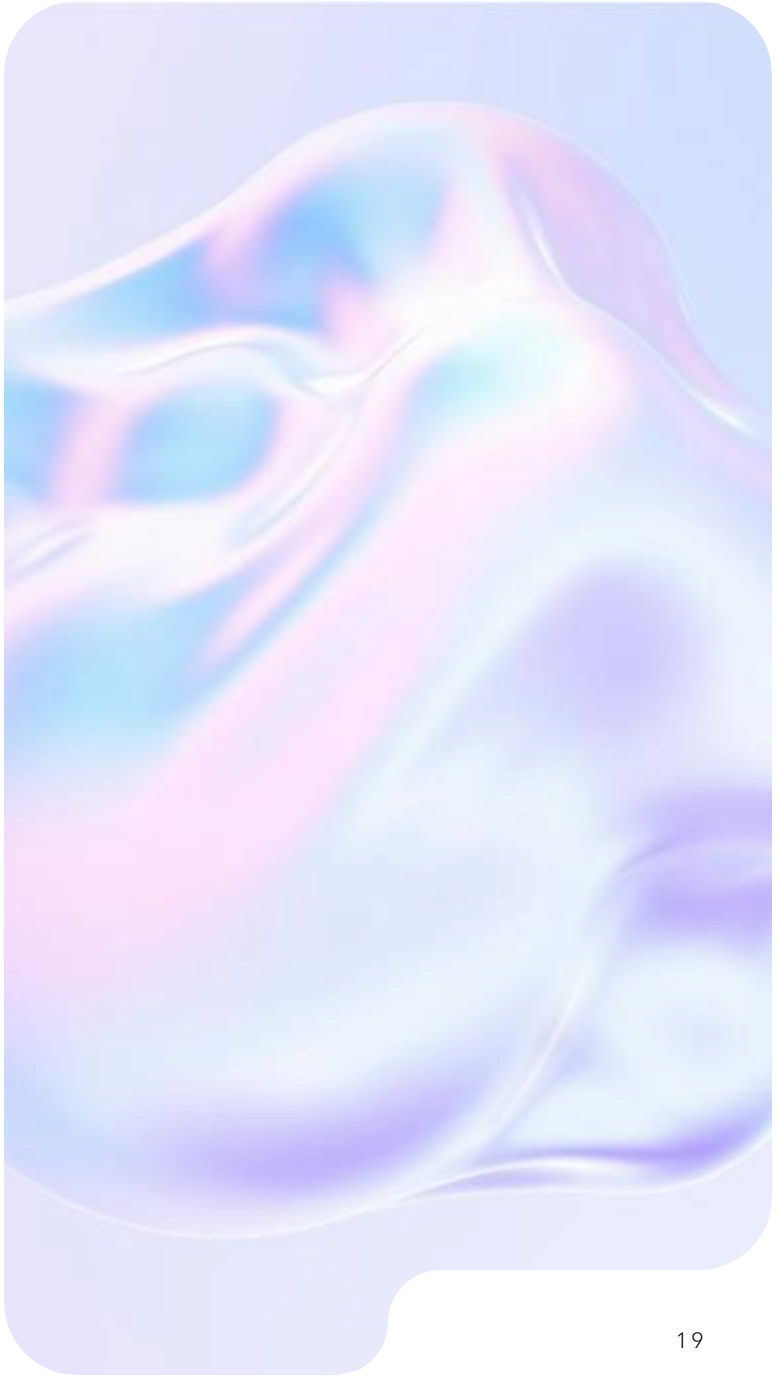
Finding data or statistics to support an argument

Outlining
- Paragraph
- Section
- Document
- Argument

Receiving feedback on drafts according to specific criteria

Brainstorming ideas

Copy editing and proofreading



Examples: Summarizing...

...an area of research

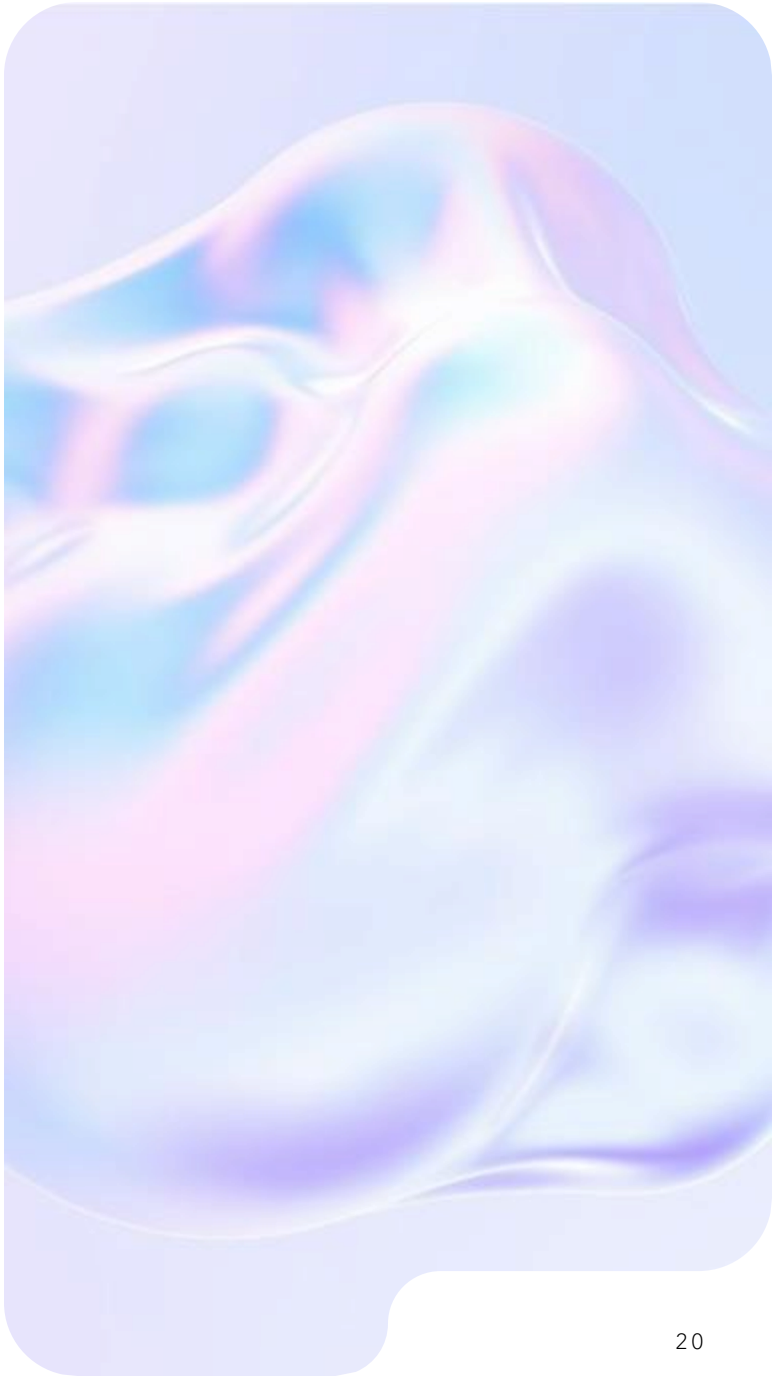
Scenario: You need to write a background paragraph explaining the "current consensus" on microglia transitioning from protective to toxic in Alzheimer's disease.

Prompt: "What is the current consensus on the transition of microglia from protective DAM to neurotoxic states in Alzheimer's disease? Return hyperlinked peer-reviewed sources in your answer."

...grant program guidelines

Scenario: You want to see if you're eligible to apply for the NIH MIRA R35 award, so you navigate to the web guidelines and ask Copilot for help.

Prompt: "On the current webpage, summarize the eligibility criteria for the NIH MIRA R35. I'm an assistant professor in the School of Biological Sciences at UC Irvine and want to see if I'm eligible."



Example: Finding statistics to support an argument

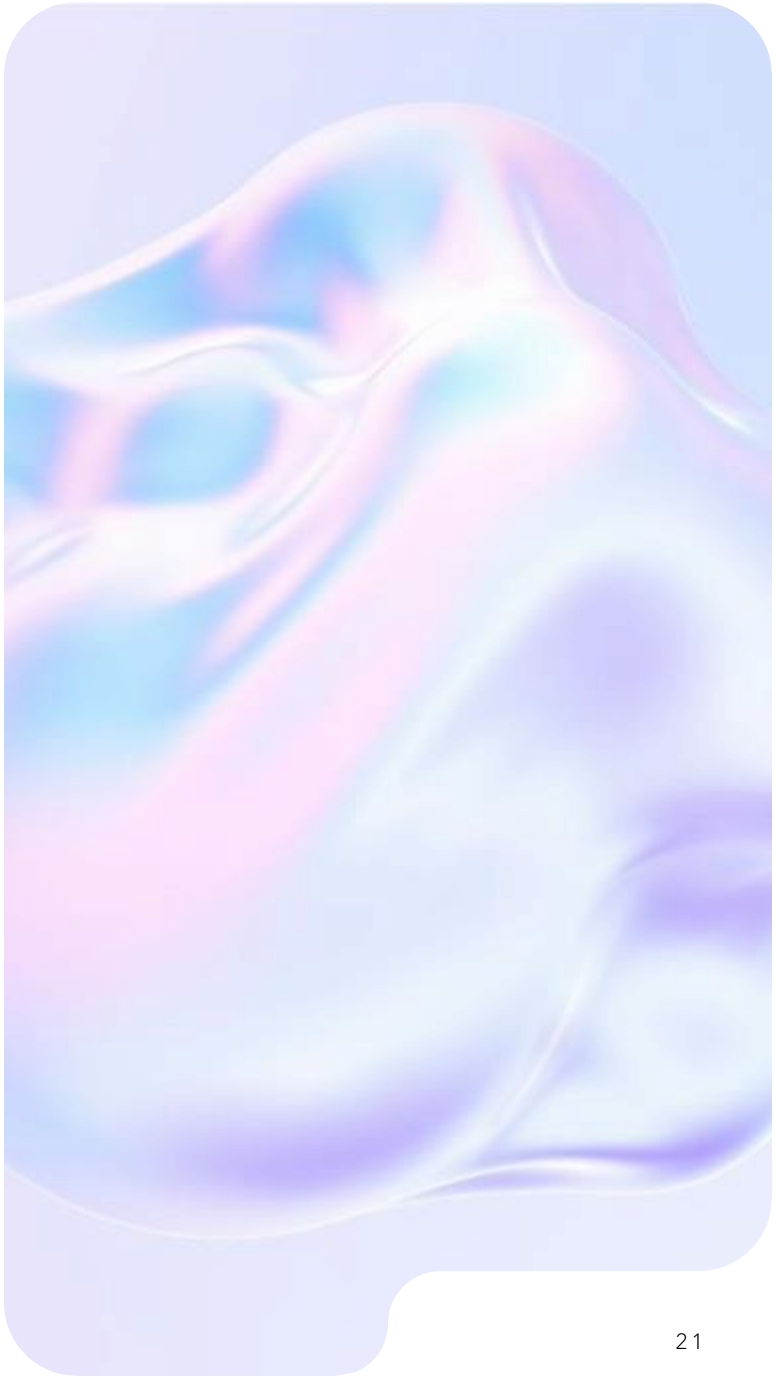
Scenario: You are testing a new method to control the invasive Indo-Pacific Lionfish. To justify the intervention, you need a "shock statistic" showing exactly how fast Lionfish consume native fish larvae to prove that passive management is failing.

Prompt: "I am writing an NSF Standard Grant about the ecological impact of invasive Lionfish on native reef fish recruitment. I need papers that quantify the 'predation rate' or 'biomass reduction.' Generate 3 complex search strings for Google Scholar that target the 'Enemy Release Hypothesis' and 'naive prey' concepts?"

Output:
"Pterois volitans" AND "recruitment limitation" AND ("native prey survival" OR "naive prey")

"Invasive lionfish" AND "trophic cascade" AND "biomass reduction" AND "mesocosm"

"Predation efficiency" AND "Pterois" AND "Atlantic vs Pacific" comparison



Example: Outlining

Scenario: You are an Associate Professor of Neuroscience applying for an R01 grant to study the role of a specific microglial pathway in early-stage Alzheimer's Disease (AD).

Challenge: You have a strong publication record, but your previous work was largely on Traumatic Brain Injury (TBI). You need to pivot your narrative to convince reviewers that your TBI expertise translates seamlessly to Alzheimer's research and that you are the right person to lead this specific project.

Goal: You want to use an LLM to outline a "Personal Statement" for your NIH Biosketch that bridges your past TBI success with your future AD research, highlighting your technical expertise in neuroinflammation as the connecting thread.

Prompt:
"I need help outlining the Personal Statement section of my NIH Biosketch for a new R01 application.

My background: 10 years studying neuroinflammation in Traumatic Brain Injury (TBI). I have developed novel mouse models for microglial activation.
The new grant: I am applying to the NIA (National Institute on Aging) to study the 'TREM2-APOE pathway in early-stage Alzheimer's.'
The problem: Reviewers might see me as just a 'TBI guy' and doubt my ability to lead an Alzheimer's project.
Task: Create a detailed outline for a Biosketch Personal Statement (approx. 1 page) that:

- Opens with a strong hook about my expertise in neuroimmune mechanisms.
- Explicitly bridges my TBI work to Alzheimer's by focusing on shared mechanisms (microglial priming).
- Highlights 4 key contributions/citations I should include (just describe what kind of papers they should be).
- Mentions my leadership and training record to reassure them I can run an R01."

Example: Receiving feedback on drafts according to specific criteria

Scenario: You've written a first draft of your Research Strategy for an NIH MIRA R35 and want to get feedback from Copilot against the NOFO's review criteria.

Prompt: "Act as an experienced NIGMS reviewer for a MIRA (R35) application under PAR-26-121. Evaluate my Research Strategy using ONLY the MIRA review criteria. For each criterion (Overall Impact, Significance, Investigator(s), Innovation, Approach, Environment), list key strengths, specific weaknesses, and whether concerns are minor/moderate/major. Assess MIRA-specific expectations (long-term vision, scope, flexibility, leadership). Flag any administrative risks (e.g., language resembling Specific Aims, excessive experimental detail). Conclude with a brief mock Overall Impact paragraph and 3–5 high-priority revision suggestions. Do NOT rewrite the text."

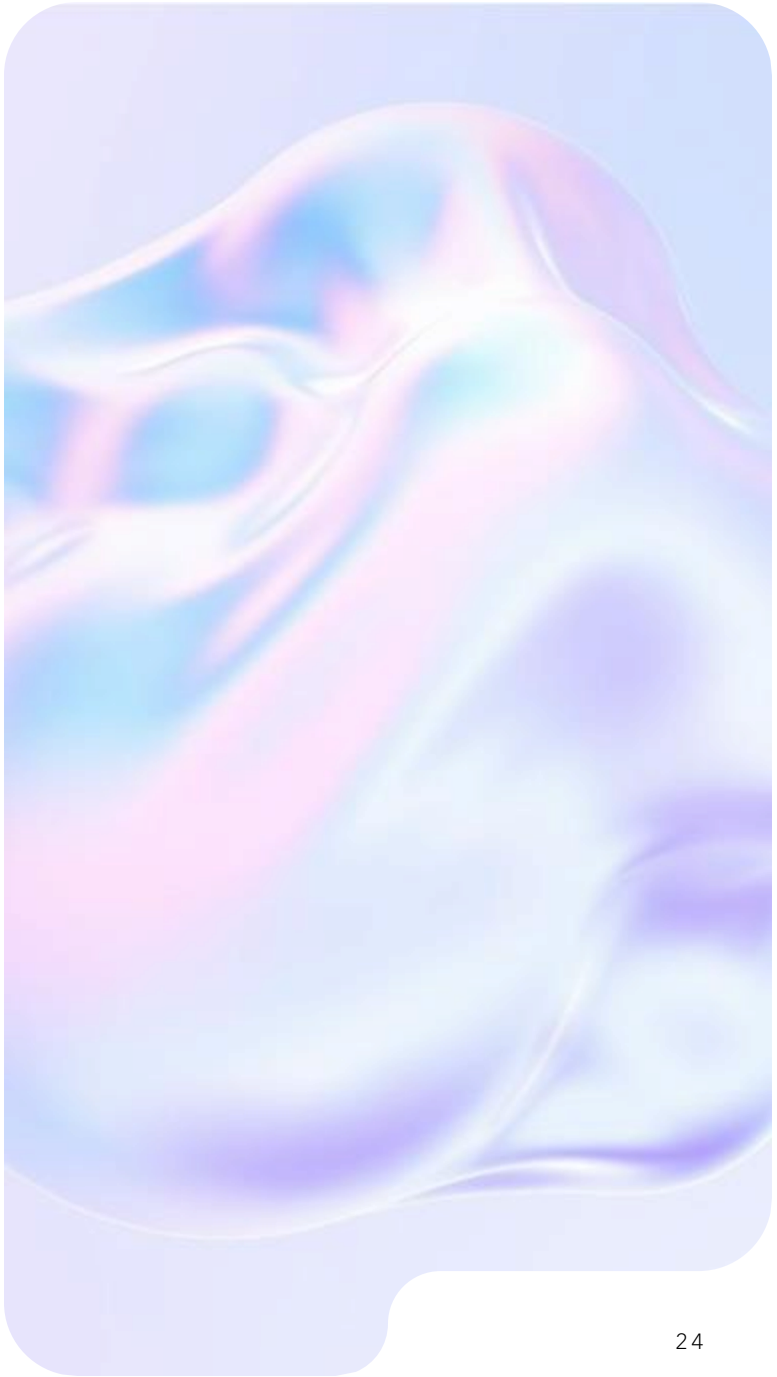
Example: Brainstorming

Scenario: You are preparing an NSF proposal on the ecological impact of invasive Indo-Pacific Lionfish on native reef fish recruitment. You need help generating strong, scientifically grounded title options that convey ecological urgency and conceptual focus.

Prompt: "Brainstorm 6 compelling NSF-style proposal titles for a project on how invasive Lionfish affect native reef fish recruitment. Titles should emphasize ecological mechanisms, recruitment impacts, and broader significance."

Output:

- "Ecological Consequences of Lionfish Invasion on Reef Fish Recruitment Dynamics"
- "Mechanisms of Recruitment Failure in Native Reef Fishes Under Lionfish Predation Pressure"
- "Invasive Predator Effects on Early-Life Survival and Community Structure in Coral Reefs"
- "Quantifying Lionfish-Driven Disruptions to Reef Fish Recovery Pathways"
- "Trophic and Recruitment-Level Impacts of Indo-Pacific Lionfish on Atlantic Reef Ecosystems"
- "Predation, Recruitment, and Reef Resilience: Assessing Lionfish Ecological Impact"



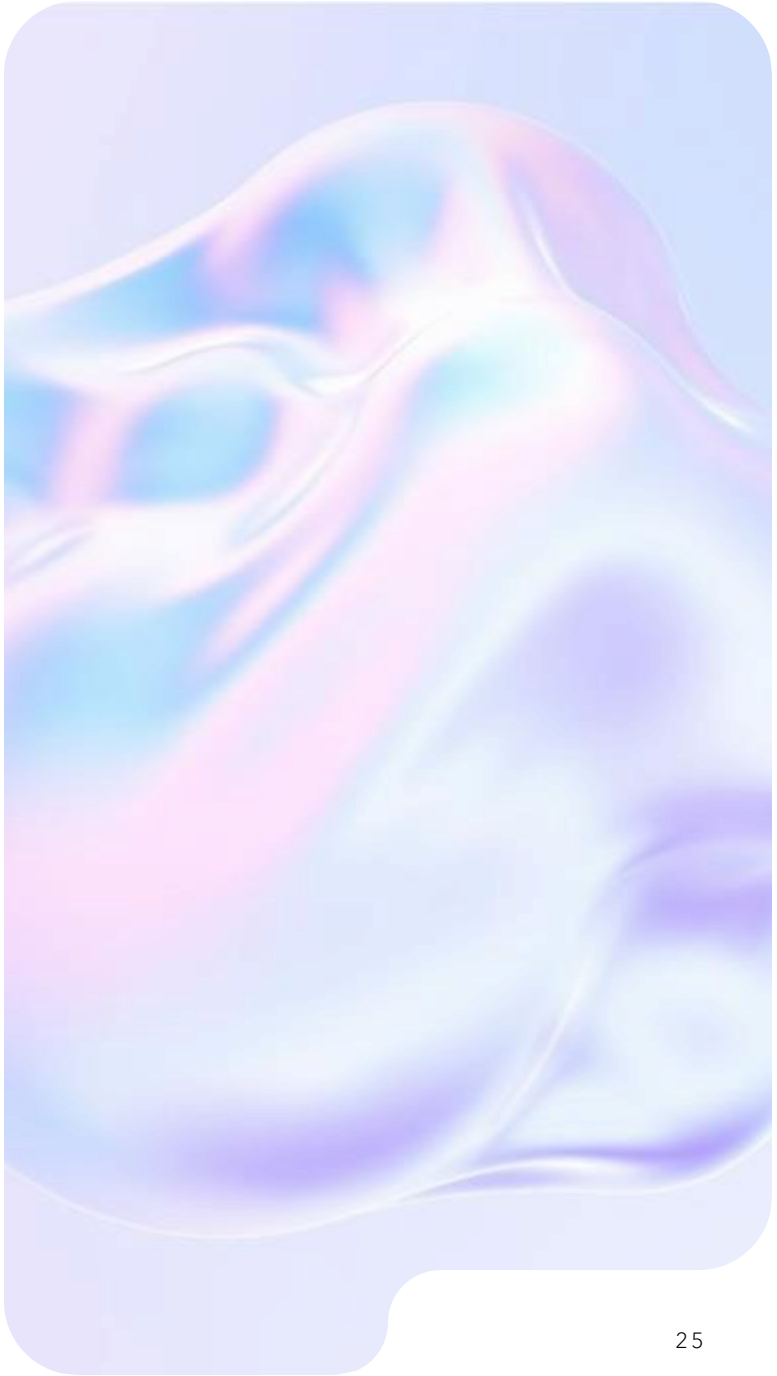
Ideas for copyediting

Use GenAI to suggest ways to improve readability, clarity, and overall flow.

Ask for recommendations on writing stronger titles and descriptive headings that act as signposts for the reader.

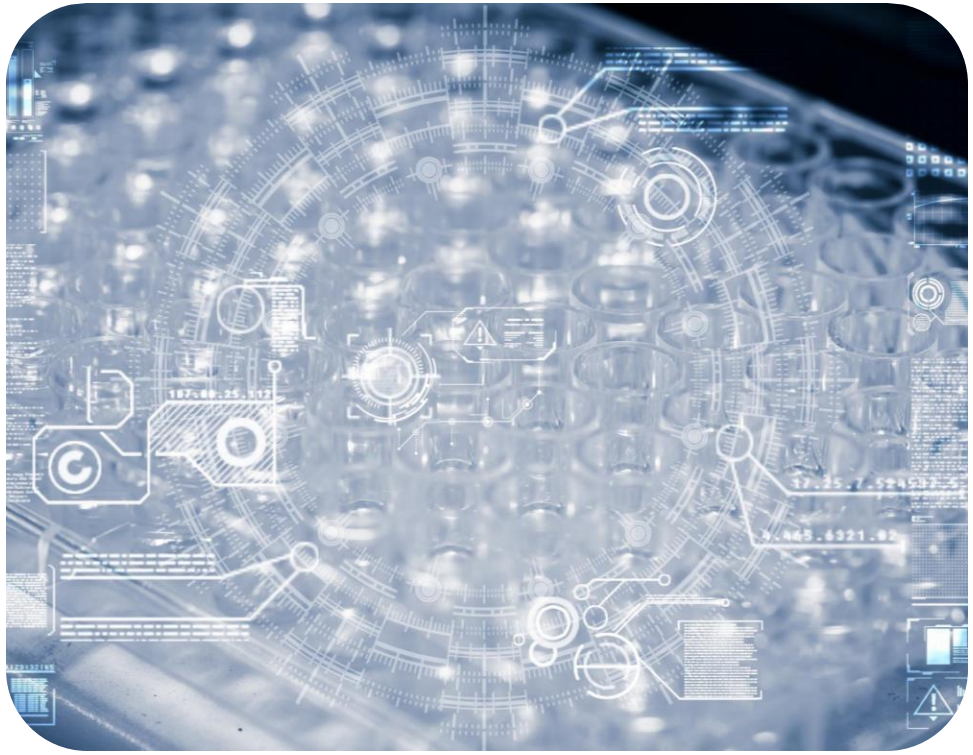
Request feedback on consistency of voice to ensure your document reads with “one voice.

Important: Instruct the GenAI not to rewrite your text. Instead, have it *suggest* specific improvements that you can implement.



Probing new (to you) research areas

Exploring New (to you) or Related Research Areas



There are many tools available to reliably probe the literature.

Literature review and conversations exploration of a topic

- Leverages GenAI's emergent ability to quickly sift through and summarize vast amounts of text.
- Some platforms allow you to interactively explore topics and supporting literature.

Augments your ability to:

- Identify research gaps
- Collaborators
- High-impact research questions

Caution

AI summaries may overgeneralize findings. Always verify summaries with original sources.

Always verify that citations are accurate.

Demonstrating practical GenAI use cases

Demo 1: Reviewing literature in a new (to you) field

Scenario:

- You're writing an NIH grant proposal on gut microbiome and neurodegeneration.
- You want to explore peripheral research on microbiome-host interactions in autoimmune diseases.
- Your goal is to identify gaps and to strengthen your Significance section.

Let's see how three AI tools could help:

Google Scholar Labs vs. PubMed GPT vs. PubMed.ai

Prompt: What are the key findings from the top 5 most recent papers on microbiome-host interactions in autoimmune diseases?

Demo 2: Brainstorming innovative proposal ideas

Google Gemini

Basic prompt:

Generate innovative research ideas for studying microbiome-host interactions.

Improved prompt:

Generate 4 innovative, high-risk/high-reward research ideas for an NIH R21 grant proposal on microbiome-host interactions.

Focus on:

- Novel mechanisms of cross-kingdom communication
- Cutting-edge single-cell or spatial omics approaches
- Therapeutic interventions with translational potential
- Understudied microbial communities or host niches
- Integration with AI/ML for pattern discovery
- Consider NIH priorities: health disparities, precision medicine, and emerging technologies
- Ideas should be feasible within 2 years with $275,000 total budget.

Demo 3: Getting feedback on drafts

Microsoft Copilot

Paragraph from a methods section:

"We performed RNA sequencing on samples collected from patients. The data was analyzed using DESeq2 and pathway analysis was done. Significant genes were identified with $p<0.05$. We also looked at immune cell populations using flow cytometry. The results showed differences between groups. Statistical analysis was performed using R."

Basic prompt:

Review this paragraph and provide feedback

Improved prompt:

Act as a peer reviewer for Nature Immunology. Review this methods paragraph for:

1. Reproducibility - Can another lab replicate this?

2. Statistical rigor - Are the methods appropriate?

3. Technical completeness - What key details are missing?

4. Field-specific standards - Does this meet current best practices for immunology research?

5. Writing clarity - How can the technical writing be improved?

Provide specific, actionable feedback with examples. Suggest precise language where appropriate. Flag any red flags that would concern reviewers.

Demo 4: Outlining a standard unit of language (sentence, paragraph, or section

Anthropic Claude Opus via ZotGPT

Prompt:

Create a structured outline for the NIH F31 Sponsor(s) Commitment section using the most recent PA-25-422 and FORMS-I instructions. Include all required headings: A. Mentoring Approach & Candidate Mentoring Plan; B. Prior Commitment to Training & Mentoring; C. Commitment to the Candidate's Research Training Plan; D. Research Training Environment; and E. Additional Institutional Commitment elements.

Context:

- F31 (PA-25-422) applicant studying microglia–synapse mechanisms in early Alzheimer's disease.
- Sponsor: 10 years of graduate mentoring; lab focuses on neuroimmune signaling.

Task:

Produce a concise outline with 3–4 bullets per required section. Do not draft narrative text. Only list what each section should cover based on the updated PA and Sponsor(s) Commitment requirements.

Inspiration

Prompting ideas for different paper sections:

Introduction: "Review for logical flow, gap identification, and hypothesis justification"

Discussion: "Evaluate interpretation balance, limitation acknowledgment, and future directions"

Abstract: "Check for compliance with 250-word structured format: Background, Methods, Results, Conclusions"

For Different Journals:

"Review as if for Science (broad audience, high impact)"

"Review as if for Journal of Biological Chemistry (technical audience, mechanistic focus)"

Prompting for scholarly writing

What makes a prompt effective?

Effective prompting = clear goal + precise communication + guided reasoning

Sound familiar? It's like explaining a task to an eager but new colleague.

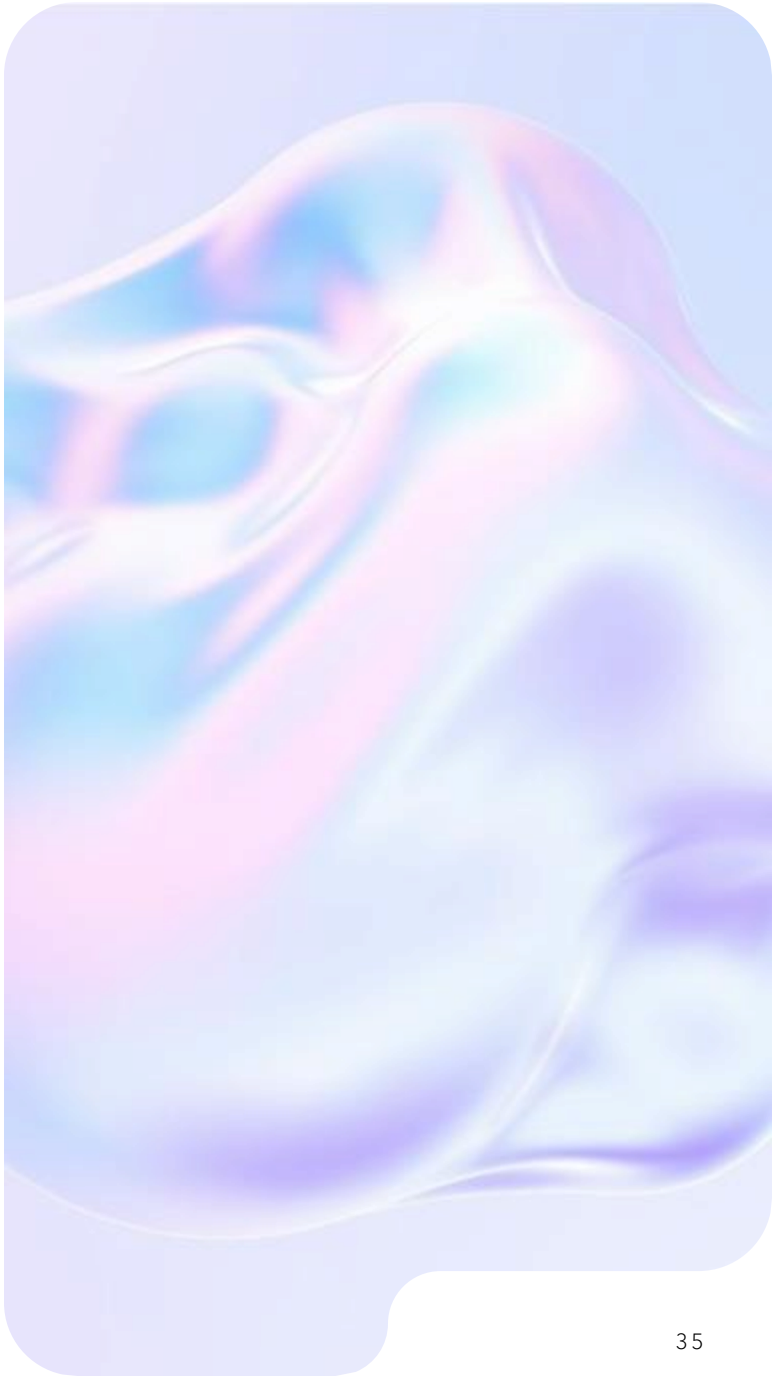
"Review this methods section."

"Review the following methods section for an NIH R01 proposal. Focus specifically on: 1. experimental design, 2. sample sizes, and 3. statistical approaches. Then, provide three recommendations to improve clarity for a reviewer unfamiliar with the field."

Effective prompting is an iterative process.

- Create a preliminary prompt
- Receive AI's response
- Refine the prompt
- Receive ideal output

Tips: Ask the AI how confident it is about its answer. Ask for different formats or variations of an output.

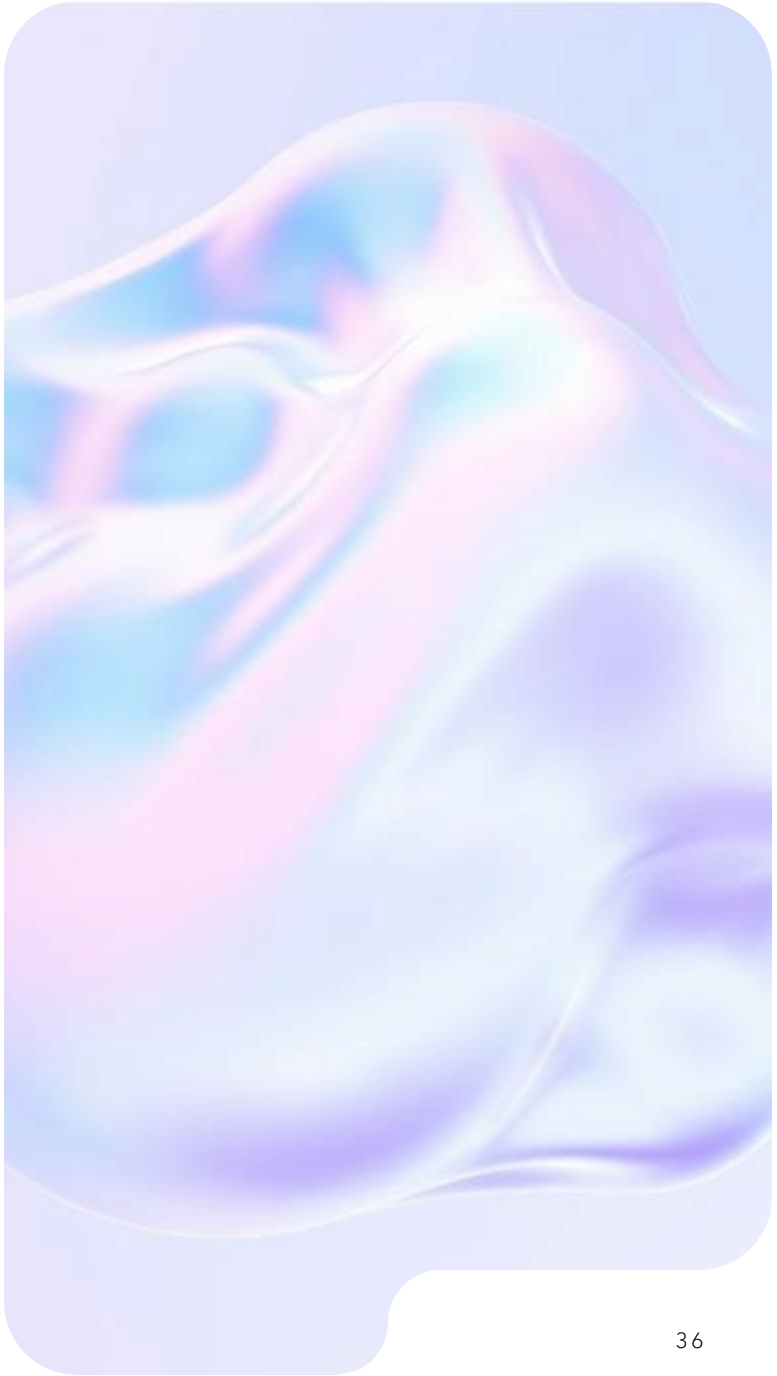


Techniques for prompting

Provide context: Be specific about what you want to know in the output as well as how you'll be using the output.

Offer examples: Use "few-shot" prompting, which means showing the AI examples of good output. This technique is good for instances when something is harder to explain than just showing examples. Try to give the AI a broad range of possible outputs.

Specify output constraints: Keep in mind variables like coding language, colors in a design, format of a response, number of bullet points, etc.



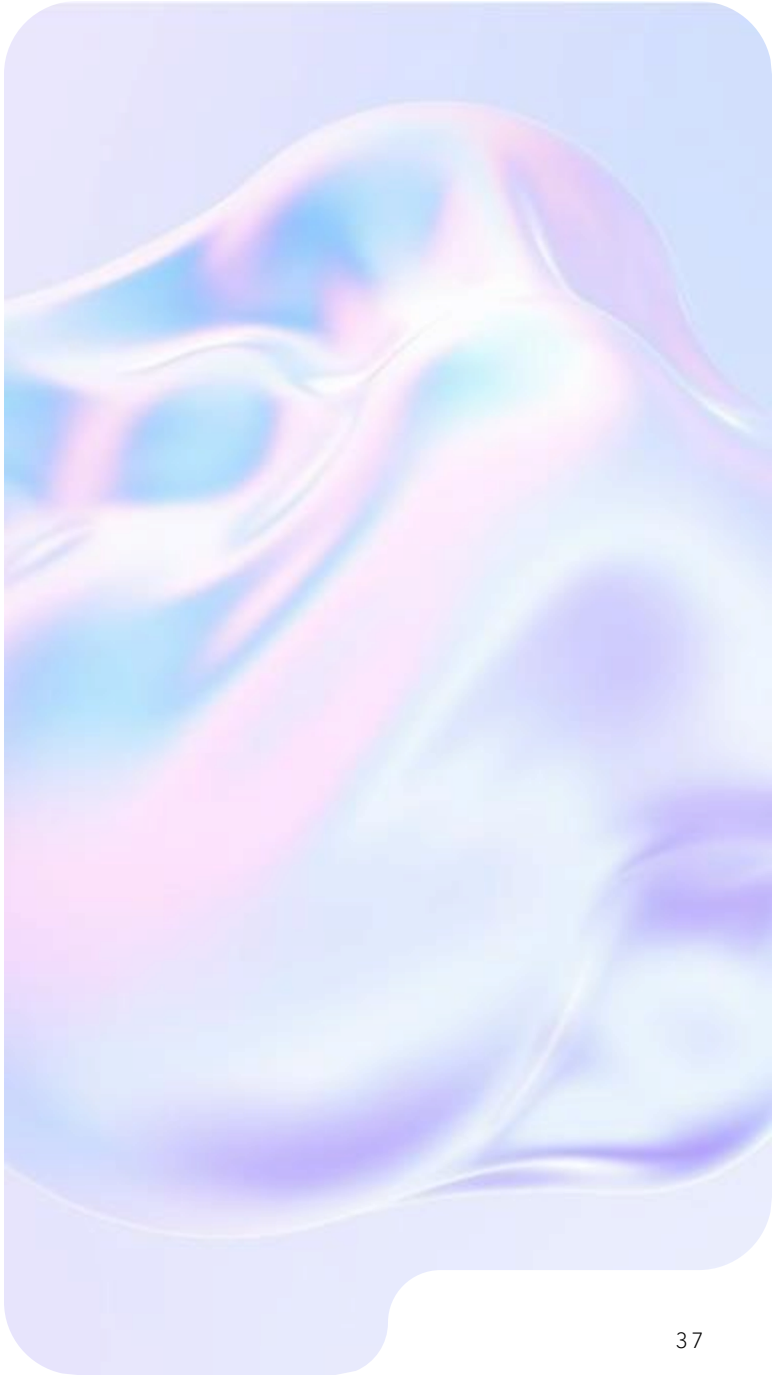
Techniques for prompting, cont'd

Give the AI space to "think": Try adding a sentence like, "Before answering, think through this problem carefully. Consider A and B before recommending the best solution."
Note: Advanced reasoning models will do this automatically, but it's important to give the AI space to think *before* it gives an answer, instead of after.

Define the role, style, or tone: This is a good way to have the AI give you different levels of explanations ("You're Richard Feynman. Explain momentum to me, a college student."). It can also be used for brainstorming and feedback ("You're an NIH grant reviewer. What do you look for when reviewing an R35 proposal?").

Pro tip: Ask the AI for help with prompting: If you're having trouble, you can ask the AI to help you phrase your request to get the best results from it.

- Use stream of consciousness to get down all of your thoughts. Then ask the AI to reformat into a prompt.



Key takeaways and responsible adoption

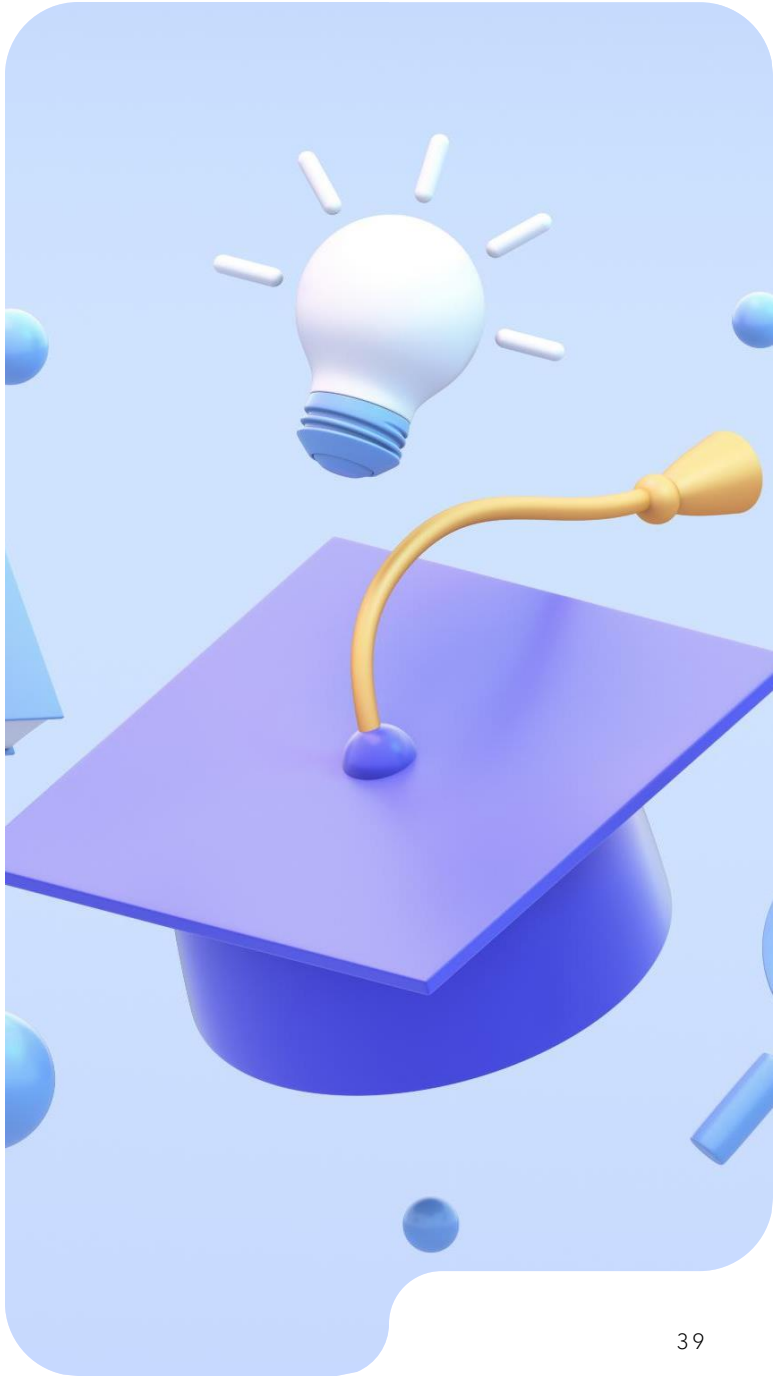
Summary of Best Practices for Faculty Use of GenAI

Use GenAI to augment—not replace—your scholarly writing. Support specific tasks like summarizing, outlining, brainstorming, and copyediting.

Avoid letting GenAI generate submission-ready text. Many funders and publishers prohibit AI-generated wording, and using it risks plagiarism and factual errors.

Maintain ownership of analysis and interpretation. GenAI cannot replace your critical thinking, disciplinary expertise, or scientific judgment.

Verify all AI-generated information. Always check claims, citations, statistics, and summaries against original sources.

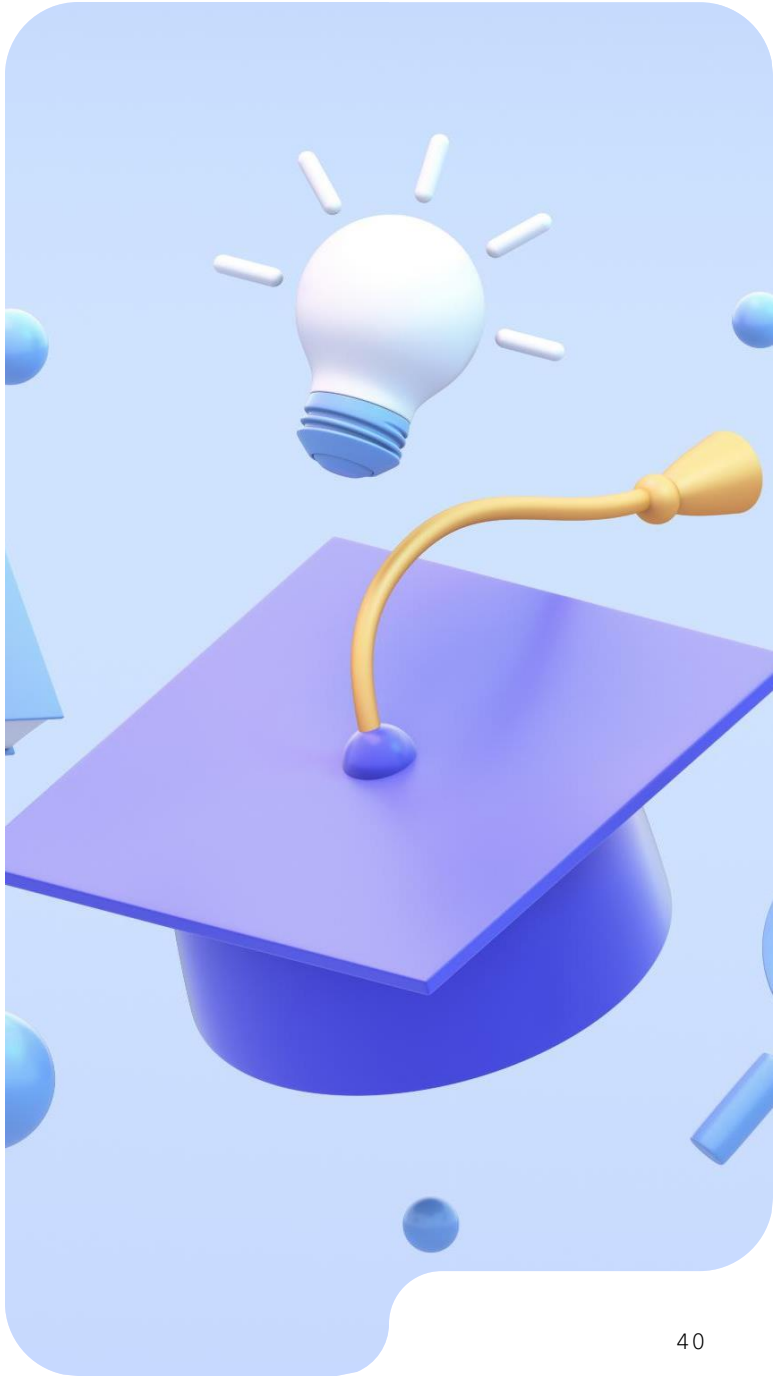


Summary of Best Practices for Faculty Use of GenAI

Use GenAI strategically to explore new research areas. Employ tools that help synthesize literature, identify gaps, and orient you to unfamiliar fields.

Prompt effectively. Provide clear goals, constraints, contextual details, and iterative refinements to get high-quality output.

Treat GenAI as an assistant. Ask for suggestions–not rewrites–when seeking clarity, readability, or consistency of voice.



Use of Generative AI in this Presentation

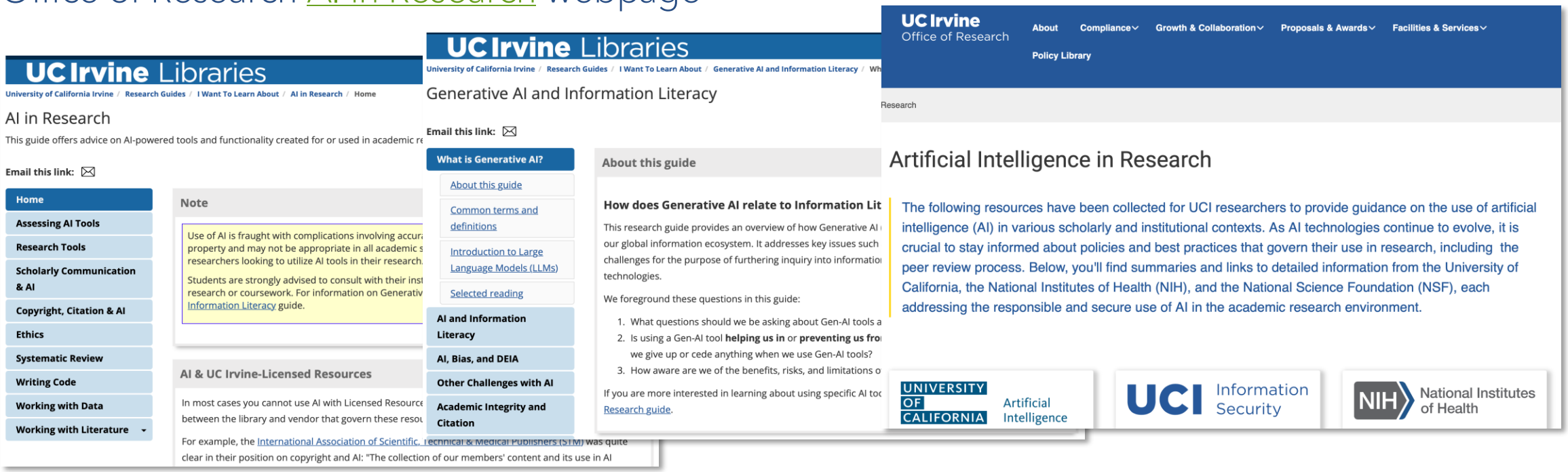
I used Google's NotebookLM to help find primary sources for the background research of this presentation. I vetted each found source and rejected those that I found not to be reliable. I then used NotebookLM to store, organize, and summarize sources. I used Microsoft's Copilot and Google's Gemini LLMs to summarize articles that I found on the internet, brainstorm examples and demo ideas, and turn the slide scripts that I wrote into bulleted talking points. I verified and approved all information presented here. The presentation content is all my own.

GenAI Ethics, Compliance & Funding Agency/Publisher Guidelines

UC Irvine Libraries

UC Irvine resources on the use of AI in research

- AI in Research library guide
- Generative AI and Information Literacy library guide
- Office of Research AI in Research webpage



Funding agency policies

Funding agency policies

- Peer review and policies on AI use
- Proposal writing and policies on AI use
- Data Management & Sharing Plans
  - Planning for AI use
  - Publishing an AI/ML model and dataset

Peer review: NIH policy on AI use

- Use of generative AI technologies is prohibited for the NIH peer review process
- "Materials pertaining to an application or proposal, and any other associated privileged information, cannot be disclosed, transmitted, or discussed with another individual through any means"…"Such actions violate NIH's peer review confidentiality requirements."
- "AI tools have no guarantee of where data are being sent, saved, viewed, or used in the future..."
- Note: Locally hosted AI platforms (e.g., ZotGPT) are explicitly prohibited (see NIH FAQs)

NOT-OD-23-49 ; NIH FAQ: Use of generative AI in peer review

Peer review: NSF policy on AI use

- "NSF reviewers are prohibited from uploading any content from proposals, review information and related records to non-approved generative AI tools."
- "Sharing proposal information with generative AI technology via the open internet violates the confidentiality and integrity principles of NSF's merit review process."
- "Any information uploaded into generative AI tools not behind NSF's firewall is considered to be entering the public domain."

Notice to Research Community: Use of Generative Artificial Intelligence Technology in the NSF Merit Review Process

Proposal writing: NIH policy on AI use

- NIH warns against using AI to write grant 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. If the detection of AI is identified post award, NIH may refer the matter to the Office of Research Integrity to determine whether there is research misconduct …"

- "if you use an AI tool to help write your application, you do so at your own risk. This is because when we receive a grant application, it is our understanding that it is the original idea proposed by the institution and their affiliated research team. Using AI tools may introduce several concerns related to research misconduct, like including plagiarized text from someone else's work or fabricated citations. If we identify plagiarized, falsified, or fabricated information in a grant write-up, we will take appropriate actions to address the non-compliance."

Proposal writing: NSF policy on AI use

- "Proposers are encouraged to indicate in the project description the extent to which, if any, generative AI technology was used and how it was used to develop their proposal."

- "Proposers are responsible for the accuracy and authenticity of their proposal submission in consideration for merit review, including content developed with the assistance of generative AI tools. NSF's PAPPG addresses research misconduct, which includes fabrication, falsification, or plagiarism in proposing or performing NSF-funded research, or in reporting results funded by NSF. Generative AI tools may create these risks, and proposers and awardees are responsible for ensuring the integrity of their proposal and reporting of research results."

Notice to Research Community: Use of Generative Artificial Intelligence Technology in the NSF Merit Review Process

NSF recent policy update

- AI tools are now included in the definition of research misconduct:
    - RESEARCH MISCONDUCT means fabrication, falsification, or plagiarism, whether committed by an individual directly or through the use or assistance of other persons, entities, or tools, including artificial intelligence (AI)-based tools

Policy Notice PAPPG 24-1, Supplement 1

Data Management & Sharing Policies

- NIH 2023 Data Sharing Policy
  - Submission of a DMS Plan with all proposals beginning January 25, 2023
  - Compliance with the DMS Plan as approved by NIH
  - Promotes broad dissemination of research products (i.e., data, research software)

- Elements to address in the DMS Plan
  - Data type
  - Related tools, software, code
  - Standards
  - Data preservation, access, timelines
  - Access, distribution, reuse considerations
  - Oversight of data management and sharing

Source: NIH Scientific Data Sharing

Publishing an AI/ML model and dataset

- Ensure digital products are FAIR (Findable, Accessible, Interoperable, Reusable)
  - Metadata
  - Code Repository
  - Dataset Card
  - Model Card
- Archive in repository and generate DOI
  - Hugging Face
  - Dryad
  - Zenodo

```
# Model Card for [Model Name]

<!-- Provide a quick summary of what the model is/does.

This model card aims to be a base template for new models. It has been generated using [this

## Model Details

### Model Description

<!-- Provide a longer summary of what this model is. -->

- **Developed by:** [More Information Needed]
- **Model type:** [More Information Needed]
- **Language(s) (NLP):** [More Information Needed]
- **License:** [More Information Needed -- choose a license (see above notes)]
- **Fine-tuned from model:** [More Information Needed]

### Model Sources

<!-- Provide the basic links for the model. -->

- **Repository:** [Project Repo]
- **Paper:** [More Information Needed--optional]
- **Demo:** [More Information Needed--encouraged]

## Uses

<!-- Address questions around how the model is intended to be used, including the foreseeable

### Direct Use

<!-- This section is for the model use without fine-tuning or plugging into a larger ecosyste
```

Guidelines, checklists, and templates available in the Collaborative Distributed Science Guide

Journal publisher policies

Writing a disclosure statement

- Disclosure statements should include:
  - Full name of the tool used (with version number)
  - How it was used
  - Reason for use

- **Statements must be included in the Methods, Acknowledgements or Appendix sections, depending on the journal**

- Authors should disclose their intent to use Generative AI tools at the earliest possible stage – either during the proposal phase or manuscript writing phase

Example disclosure statements

- "The authors acknowledge the use of Microsoft Copilot (accessed August 2024) to generate a first draft of the Acknowledgments section and translate the Methods section from Spanish to English. Gemini (accessed July 2025) was used to reorganize the Discussion section for clarity and flow. All AI-assisted text was reviewed and revised by the authors to ensure accuracy and clarity of meaning."

- "Literature screening was assisted by GPT-4 (accessed March 2024) to categorize 500+ abstracts by predefined relevance criteria. AI-assisted outputs were reviewed and validated by the authors, who confirmed the final inclusion and exclusion decisions."

Maintaining records of AI assistance

"The final products, publications, lecture materials and e-mails that had already been exported or used still exist. What disappeared, however, was my entire underlying correspondence with ChatGPT: the prompts, follow-up questions, iterations and project folders. This process history was the foundation for developing the outputs. *Losing it was like losing the laboratory notebooks behind published research – the end results remain, but the developmental record and reusable building blocks are gone.*"

Maintaining records of AI assistance

- It is important to maintain clear records of your AI use throughout your research – this will be helpful when adapting disclosures as needed

- Recommendations on maintaining records:
  - Preserve the output of AI in case it is requested during peer review or after publication
  - Document the prompts used with AI tools
  - Create independent backups of all AI-assisted work
  - Note the AI tool and version used, month accessed, and why it was used

APA Journals policy on generative AI; Using AI tools in your research: A guide for research authors, editors, and reviewers

Journal policies and submission guidelines

Journal of Biological Chemistry

- "We now require that the use of such services should be transparently and specifically declared in the **acknowledgment section of submitted manuscripts**. Importantly, it is inappropriate for ChatGPT or similar technologies, or more traditional commercial editing services, to be included as a co-author, because such technologies/services cannot meet all of the essential authorship requirements."

- "… confidentiality is a core principle of the peer-review process… Thus, uploading any part of the manuscript into a public AI tool may violate the authors' confidentiality and proprietary and/or data privacy rights, and manuscripts should not be evaluated using generative AI tools."

Journal policies and submission guidelines

Brain Research

- "The use of AI Tools in the manuscript preparation process must be declared by adding a statement at the end of the manuscript when the paper is first submitted. The statement will appear in the **published work** and should be placed in a **new section before the references list**."

- Example:
  - Title of new section: Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.
  - Statement: During the preparation of this work the author(s) used [NAME OF TOOL / SERVICE] in order to [REASON]. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Journal policies and submission guidelines

Scientific Reports

- "Use of an LLM should be properly documented in the **Methods section** (and if a Methods section is not available, in a suitable alternative part) of the manuscript."
- "The use of an LLM (or other AI-tool) for "AI assisted copy editing" purposes does not need to be declared."
- "In all cases, there must be human accountability for the final version of the text and agreement from the authors that the edits reflect their original work."
- "While legal issues relating to AI-generated images and videos remain broadly unresolved, **Springer Nature journals are unable to permit its use for publication**."

Journal policies and submission guidelines

Environmental Science and Technology

- "The use of AI tools for text or image generation should be disclosed in the manuscript within the **Acknowledgment section with a description of when and how the tools were used.** For more substantial use cases or descriptions of AI tool use, authors should provide full details within the Methods or other appropriate section of the manuscript."

- "For graphics, a brief description of AI use should be included in the figure caption to explain to readers how the image was created."

- "Disclosing any part of the submission or the peer review report itself to a text generation service (or any other third-party tool) is a breach of this confidentiality, and thus a violation of the ACS Ethical Guidelines."

How to navigate conflicting AI guidelines

- Review UC Irvine's AI policy and note the date when you last checked it
- Check for specific requirements of a journal, as policies may vary across publishers and titles
- Document your AI use consistently throughout your research
- When in doubt, follow the most restrictive policy as that is the safest approach
- Always comply with the terms of your author agreement concerning AI and intellectual property

Thank you!

- Tiffany Esteban, testeba1@uci.edu
- Wasila Dahdul, wdahdul@uci.edu
- Angelissa Panganiban, angelissa.panganiban@uci.edu

UC Irvine Libraries

Using AI to extend expertise

CHRISTOPHER C. W. HUGHES

CHANCELLOR'S PROFESSOR

MOLECULAR BIOLOGY & BIOCHEMISTRY

The problem

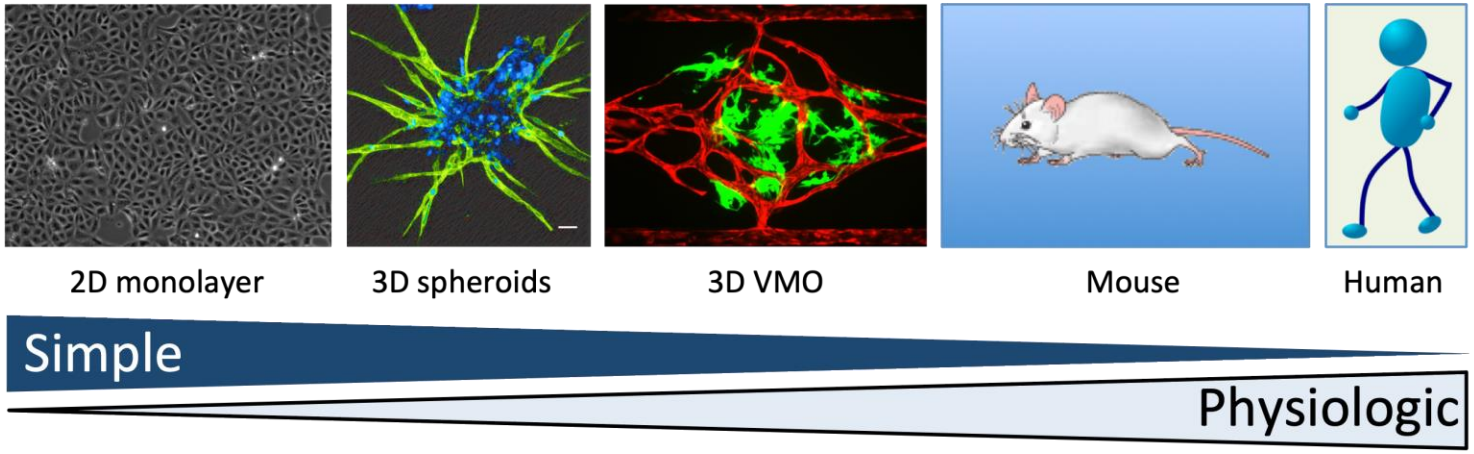
My expertise:

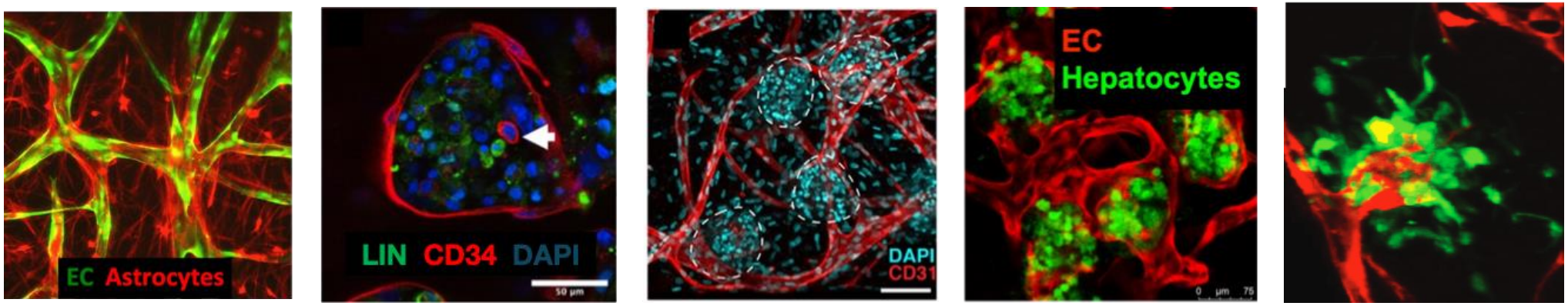
Cell and molecular biology, with a “minor” in biomedical engineering

Where I need to go:

Surface chemistry

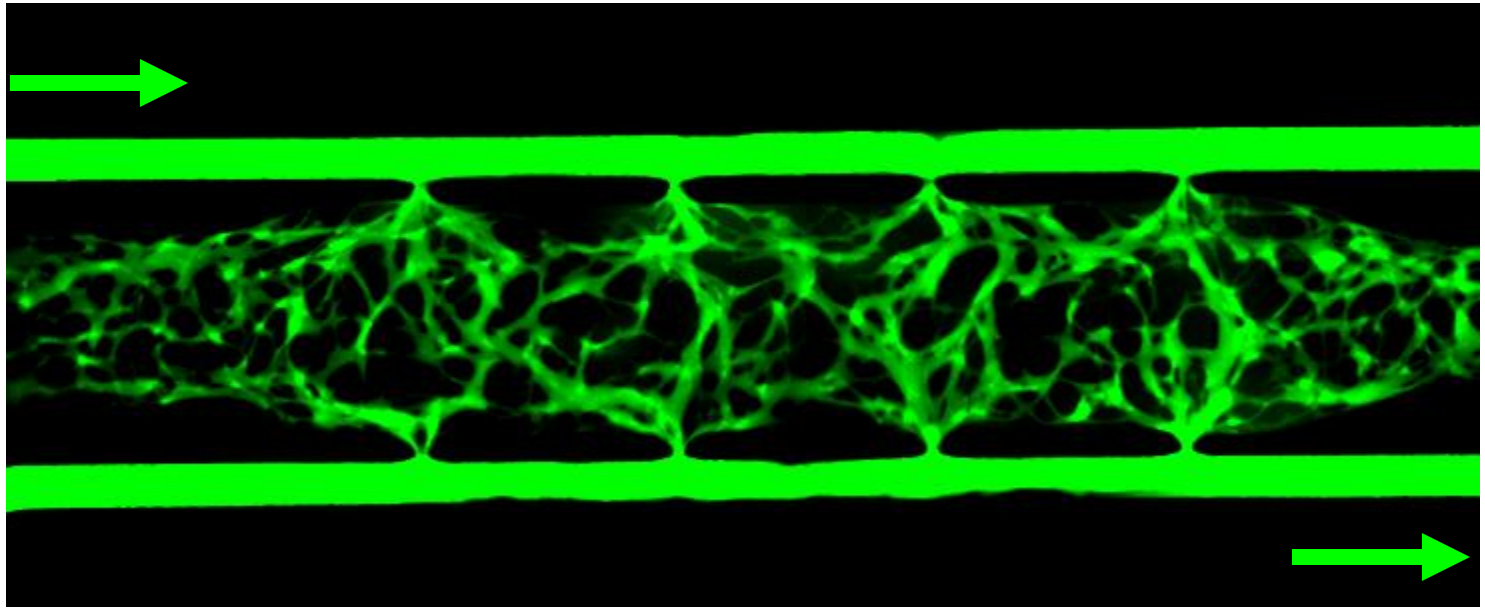
What we do – microphysiological systems



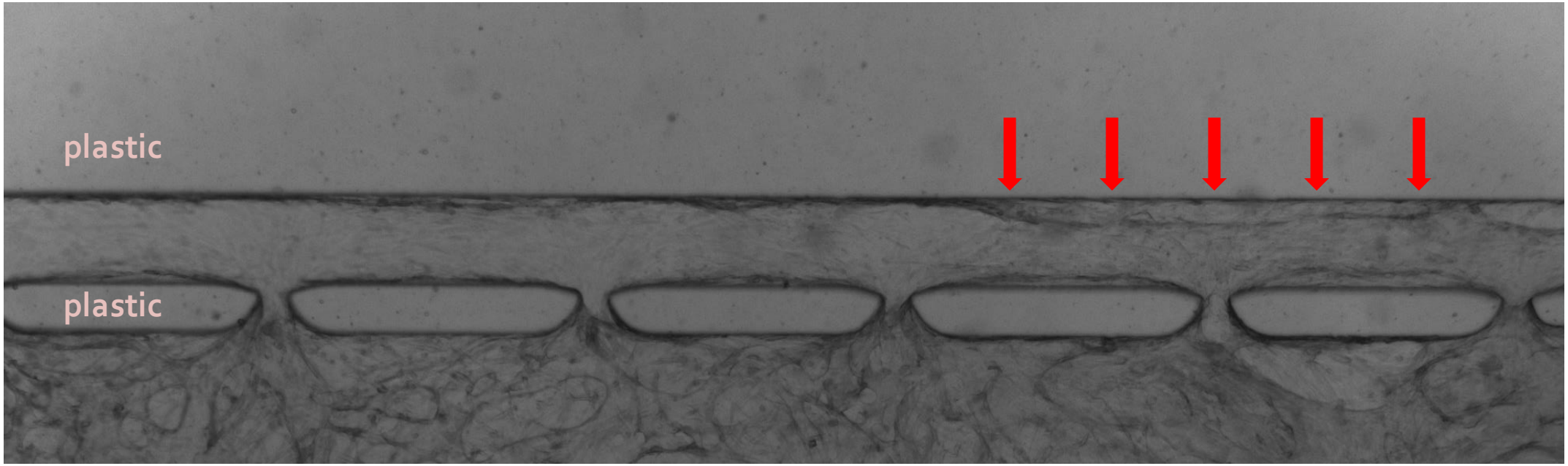


Brain | Bone Marrow | Pancreas | Liver | Patient tumor

All have blood vessels

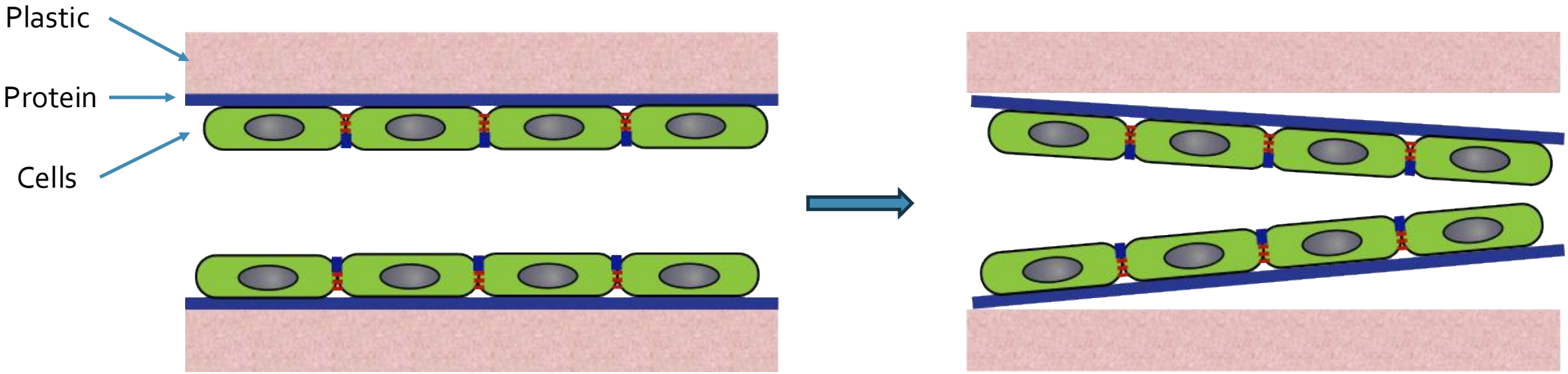


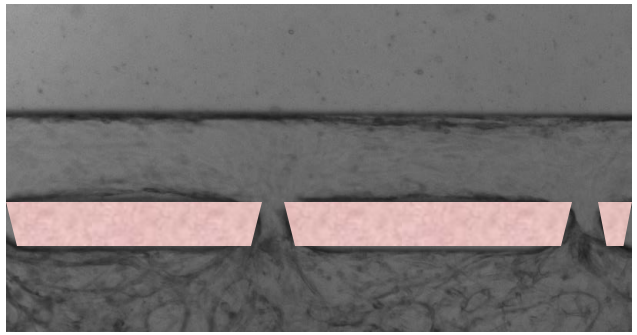
But, cells peel off the outer channels…

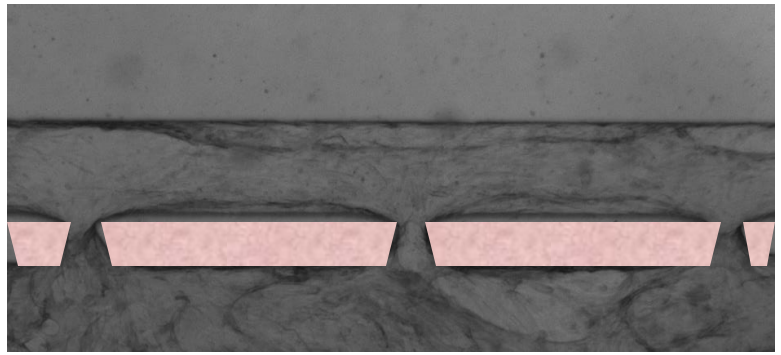


… vessels collapse, flow stops, cells die

Protein basement membrane detaches







How do you stick (covalently attach) protein to plastic?

Ask your colleagues in Chemistry!

“yeah, sorry mate, not my field…”

How do you stick (covalently attach) protein to plastic? Part 2.

Ask ZotGPT!

Specifically…

ZotGPT, Anthropic Claude Sonnet 3.7 – 4.0 (Balanced/Model)

You

is there a way to modify proteins so that they will covalently attach to the surface of pdms?

3/7/2025 2:24 pm

ZotGPT

Yes, there are several approaches to modify proteins for covalent attachment to polydimethylsiloxane (PDMS) surfaces. Here are a few methods:

1. Silane coupling:
   - Modify the PDMS surface with silane coupling agents containing reactive groups (e.g., epoxy, amine, or thiol groups).
   - Modify proteins with complementary reactive groups that can form covalent bonds with the silane-modified PDMS surface.
2. UV-activated crosslinking:
   - Incorporate photoreactive groups (e.g., benzophenone derivatives) into proteins.
   - Upon UV irradiation, these groups can form covalent bonds with the PDMS surface.
3. Click chemistry:
   - Modify proteins with azide or alkyne groups.
   - Functionalize the PDMS surface with complementary alkyne or azide groups.
   - Use copper-catalyzed azide-alkyne cycloaddition (CuAAC) to form covalent bonds.
4. Plasma treatment and EDC/NHS chemistry:
   - Treat PDMS surface with oxygen plasma to introduce hydroxyl groups.
   - Activate these groups with EDC/NHS chemistry.
   - React with amine groups on proteins to form amide bonds.
5. Thiol-ene chemistry:
   - Modify proteins to introduce thiol groups.
   - Functionalize PDMS with alkene groups.
   - Use UV light or a radical initiator to catalyze the thiol-ene reaction.

Each method has its advantages and limitations, and the choice depends on the specific protein, desired orientation, and application requirements.

3/7/2025 2:25 pm

the binding cannot involve non-biocompatible reagents or UV

3/7/2025 2:25 pm

You

ZotGPT

Understood. Given the constraints of biocompatibility and avoiding UV light, here are some protein modification approaches that could potentially work for covalent attachment to PDMS surfaces:

1. Enzymatic coupling:
   - Modify proteins with specific peptide tags (e.g., ybbR tag or SpyTag).
   - Use enzymes like Sfp phosphopantetheinyl transferase or SpyCatcher to catalyze the formation of covalent bonds between the tagged proteins and appropriately functionalized PDMS surfaces.
2. Sortase-mediated ligation:
   - Engineer proteins with a C-terminal LPXTG motif.
   - Modify the PDMS surface with oligoglycine.
   - Use sortase A to catalyze the transpeptidation reaction, forming a covalent bond between the protein and the surface.
3. Native Chemical Ligation (NCL):
   - Introduce an N-terminal cysteine residue to the protein of interest.
   - Modify the PDMS surface with thioester groups.
   - The NCL reaction will form a native peptide bond between the protein and the surface.
4. Oxidative coupling:
   - Introduce tyrosine residues at specific locations in the protein.
   - Use mild oxidizing conditions (e.g., horseradish peroxidase and hydrogen peroxide) to create dityrosine crosslinks between the protein and tyrosine-functionalized PDMS surfaces.
5. Bioorthogonal chemistry:
   - Incorporate unnatural amino acids with bioorthogonal reactive groups (e.g., azides or strained alkynes) into proteins using genetic code expansion techniques.
   - Modify PDMS surfaces with complementary reactive groups.
   - Use strain-promoted azide-alkyne cycloaddition (SPAAC) or other bioorthogonal reactions to form covalent bonds.

Modifications must be biocompatible, not use UV and must generate covalent bonds

3/7/2025 3:48 pm

You

ZotGPT

Thank you for
focus on the r

1. Silane cou
   - Modify
   - These g
   - This me
   - Process
     peptide
2. EDC/NHS c
   - This me
   - Process
     carboxy
     peptide
3. Dopamine
   - This me
   - Process
     peptide
4. Epoxy-bas
   - This me
   - Process
     hydroxy
5. Maleimide
   - This ap
   - Process
     maleim

Among these

- The specifi
- The desire
- The stabili
- The availa

will the plasma-treated surface of PDMS covalently bond proteins?

3/7/2025 2:40 pm

You

ZotGPT

Plasma-treated Pl
Here's a more det

1. Surface modifi
   - Plasma trea
   - These group
2. Interaction with
   - The plasma
   - Instead, it pr
3. Potential for co
   - In some cas
     is not the pri
4. Enhanced prot
   - The hydroph
5. Additional step
   - To ensure c
   - For example
     covalent bo
6. Time sensitivity
   - Due to the s
7. Variability:
   - The exact n
     environmen

In summary, while
For assured coval
using the plasma

3/7/2025 2:41 pm

is there a specific laminin peptide that endothelial cells bind to?

3/7/2025 2:52 pm

You

ZotGPT

Yes, there are specific
identified as particular

1. YIGSR (Tyr-Ile-Gly-S
   - This peptide, der
   - It's recognized b
2. IKVAV (Ile-Lys-Val-A
   - Found in the α1 c
   - It's also involved
3. AG73 (RKRLQVQLSIR
   - This 12-amino ac
   - It binds to synde
4. C16 (KAFDITYVRLKF):
   - Derived from the
   - It interacts with i
5. A13 (RQVFQVAYIIIKA
   - This peptide fron
6. RGD (Arg-Gly-Asp):
   - While not specifi
     their adhesion.
7. REDV (Arg-Glu-Asp
   - This sequence, fo

Key points to consider:

- The effectiveness of
- Often, combination
- The 3D presentation
- Some of these pept
  engineering applica

how can peptides be modified so that they bind to PdMS?

3/7/2025 3:47 pm

You

ZotGPT

To modify peptides for bind
the modifications typically

1. Silane coupling:
   - Modify peptides with
   - These groups can for
   - Reference: Kuddanna
     mesenchymal stem c
2. Hydrophobic modificatio
   - Add hydrophobic am
   - This can increase nor
   - Example: Adding a po
   - Reference: Li, Y., et al.
3. Plasma treatment follow
   - Treat PDMS with oxyg
   - Modify peptides with
   - Use EDC/NHS chemist
   - Reference: Tan, J. L., e
     Engineering Part C: M
4. Click chemistry:
   - Modify PDMS with azid
   - Use copper-catalyze
   - Reference: Liang, R. P.,
     Analytical Chemistry,
5. Photochemical immobili
   - Modify peptides with
   - Upon UV irradiation, t
   - Reference: Toh, C. R., e
     microfluidic device ap
6. Dopamine-based modif
   - Coat PDMS with polyd
   - Couple peptides to th
   - Reference: Lee, H., et a

Is there a compound you could mix with the protein to promote covalent bonding to plasma-treated pdms?

3/7/2025 2:42 pm

You

ZotGPT

Yes, there are compounds you can mix with proteins to promote covalent bonding to plasma-treated PDMS. One of the most commonly used and effective approaches is to use a crosslinking agent. Here are some options:

1. Glutaraldehyde:
   - A widely used crosslinking agent that can react with amine groups on proteins and hydroxyl groups on the plasma-treated PDMS surface.
   - Pros: Highly effective, readily available
   - Cons: Can potentially alter protein structure or function
2. Carbodiimides (e.g., EDC):
   - These compounds can activate carboxyl groups on proteins to react with amine or hydroxyl groups on the surface.
   - Often used in combination with N-hydroxysuccinimide (NHS) for improved efficiency.
   - Pros: Mild reaction conditions, water-soluble
   - Cons: Short half-life in aqueous solutions
3. BS3 (Bis(sulfosuccinimidyl)suberate):
   - A homobifunctional crosslinker that reacts with primary amines on proteins and surface.
   - Pros: Water-soluble, non-cleavable
   - Cons: May cause protein-protein crosslinking
4. Sulfo-SANPAH (Sulfosuccinimidyl 6-(4'-azido-2'-nitrophenylamino)hexanoate):
   - A heterobifunctional crosslinker that can be activated by visible light.
   - One end reacts with amines on proteins, the other with the PDMS surface.
   - Pros: Allows for spatial control of protein attachment
   - Cons: Light-sensitive, requires careful handling
5. SMCC (Succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate):
   - A heterobifunctional crosslinker that reacts with amines and sulfhydryl groups.
   - Can be used if your protein has accessible cysteine residues.
   - Pros: Highly specific, forms stable bonds
   - Cons: May require protein modification if no accessible cysteines are present

When using these compounds:

- Mix the chosen crosslinker with your protein solution before applying to the plasma-treated PDMS surface.
- Follow the specific reaction conditions (pH, temperature, time) recommended for each crosslinker.
- Be aware that crosslinking may affect protein structure or function, so optimization may be necessary.
- Consider the biocompatibility of the crosslinker if it's crucial for your app[illegible]ion.

STARK LAB

provide references on the best combination of laminin peptides to promote endothelial cell adhesion

3/7/2025 2:54 pm

You

ZotGPT

Certainly! Here are some key references that discuss combinations of laminin peptides to promote endothelial cell adhesion:

1. Por… Nomizu, M., Delgado, M. C., Kuratomi, Y., Hoffman, M. P., Powell, S., … & Kleinman, H. K. (1999). Identification of endothelial cell binding sites on the …γ1 chain. Circulation research, 84(6), 688-694.
   - …y identifies multiple active sites on the laminin γ1 chain and shows that combinations of these sites are more effective than individual peptides.
2. Kid… Williams, S. K. (2004). Laminin-5-enriched extracellular matrix accelerates angiogenesis and neovascularization in association with ePTFE. Jou… …omedical Materials Research Part A: An Official Journal of The Society for Biomaterials, 69(2), 294-304.
   - …r demonstrates the effectiveness of laminin-5, which contains multiple bioactive sequences, in promoting endothelial cell adhesion and angiogenesis.
3. Ma… …l., Nomizu, M., Chung, M., Delgado, M., Kuratomi, Y., Yamada, Y., … & Kleinman, H. K. (1999). Identification of laminin α1 and β1 chain peptides act… …dothelial cell adhesion, tube formation, and aortic sprouting. The FASEB Journal, 13(1), 53-62.
   - This study identifies multiple peptides from laminin α1 and β1 chains that promote endothelial cell adhesion and angiogenesis, suggesting potential combinations.
4. Kikk… …ujiwara, T., Sudo, R., Sekiguchi, K., Taniguchi, M., & Mochizuki, N. (2017). Laminin-111 and laminin-511 promote endothelial cell adhesion and fun… …entific reports, 7(1), 1-11.
   - This paper compares the effects of different laminin isoforms on endothelial cells, providing insights into the most effective combinations of laminin sequences.
5. Sagnella, S., Anderson, E., Sanabria, N., Marchant, R. E., & Kottke-Marchant, K. (2005). Human endothelial cell interaction with biomimetic surfactant polymers containing Peptide ligands from the heparin binding domain of fibronectin. Tissue engineering, 11(1-2), 226-236.
   - While focusing on fibronectin, this study demonstrates the effectiveness of combining multiple peptide sequences for improved endothelial cell adhesion.
6. Le S… …lagenau, A., Gunaratnam, K., Kilian, K. A., Böcking, T., Gooding, J. J., & Gaus, K. (2011). Spacing of integrin ligands influences signal transduction in e… …l cells. Biophysical journal, 101(4), 764-773.
   - …r highlights the importance of peptide spacing and presentation in addition to sequence combination for optimal endothelial cell adhesion.
7. Yar… …M., Yamamoto, K., & Noumura, T. (1993). Type I collagen promotes modulation of cultured rabbit arterial smooth muscle cells from a con… …o a synthetic phenotype. Experimental cell research, 204(1), 121-129.
   - While not specifically about laminin peptides, this study demonstrates the synergistic effects of combining different ECM components on vascular cell behavior.

BS

BS

BS

BS

BS

Citations are a complete fantasy!

Titles don't exist

References are to other, unrelated papers

However, *information* is mostly solid

STARK LAB

Outcome

Ran the ideas by a chemist – sounds good!

First test – promising, but some issues

Back to ZotGPT…

Outcome

Ran the ideas by a chemist – sounds good!

First test – promising, but some issues

Back to ZotGPT…

Second test – much improved, but new issue

Back to ZotGPT…

Outcome

Ran the ideas by a chemist – sounds good!

First test – promising, but some issues

Back to ZotGPT…

Second test – much improved, but new issue

Back to ZotGPT…

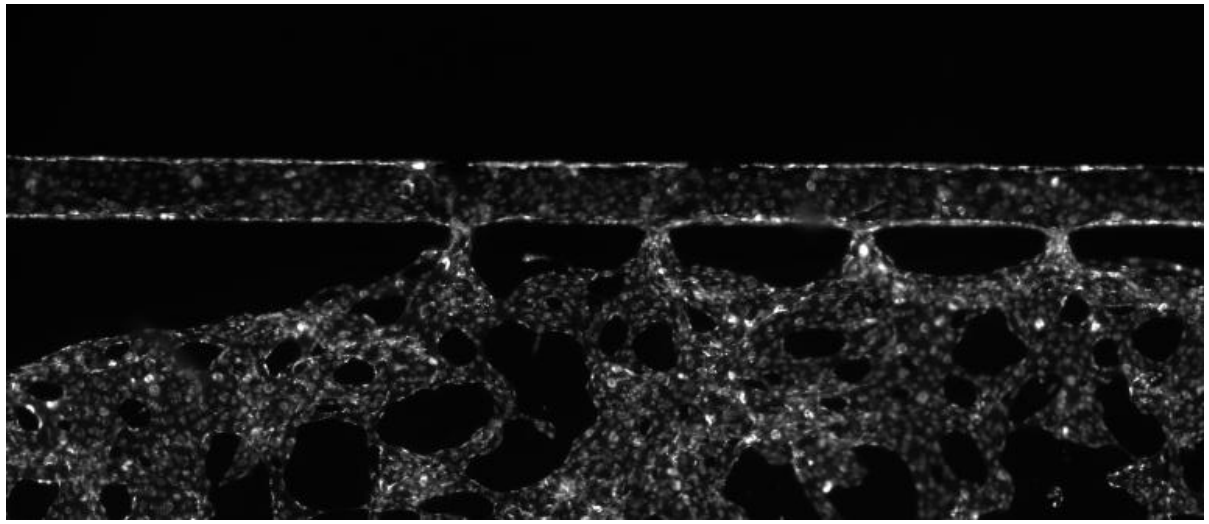
Third test – almost there…

Back to ZotGPT

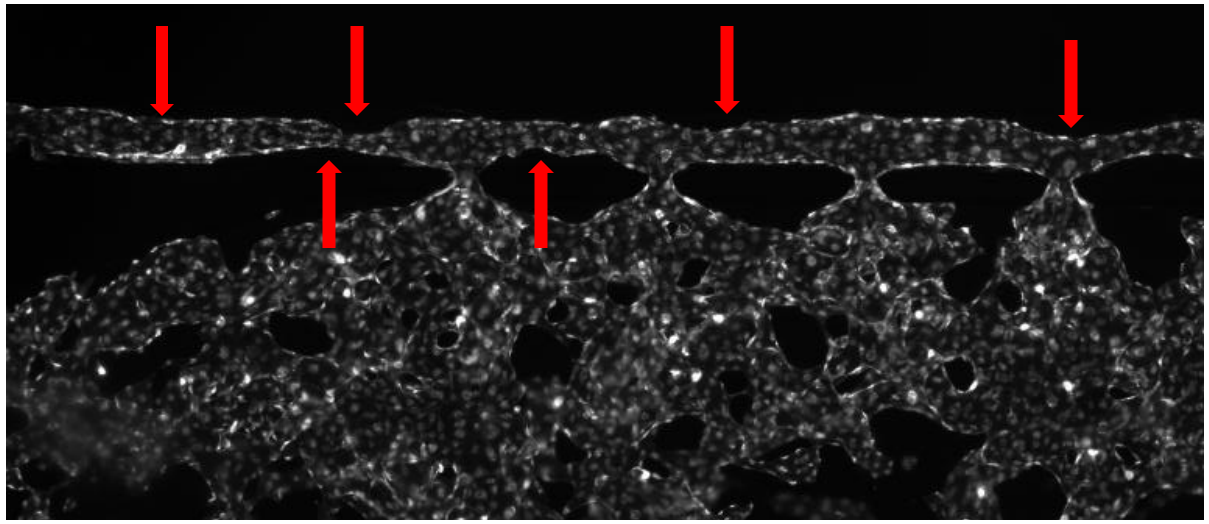
However…

It works!

New method



Old method



Recommendations

- It’s an iterative process
- Ask natural questions
- Follow up with detailed questions
- Challenge
- Go back with failures, ask for solutions
- Do not assume you get everything “Claude” knows in the first response
- Do. Not. Trust. Citations.

Don’t be afraid to explore new fields!

Thank you – questions?

Learning Single Cells using AI

Qing Nie

Department of Mathematics

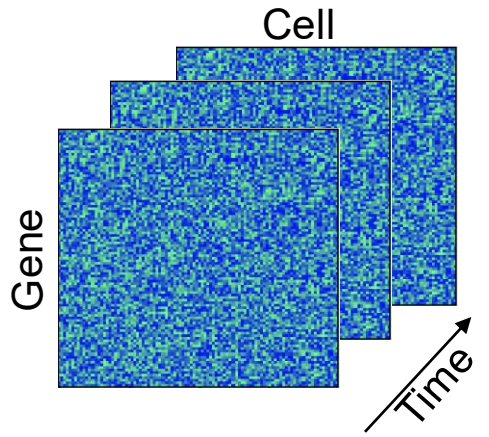
Department of Developmental and Cell Biology

University of California, Irvine

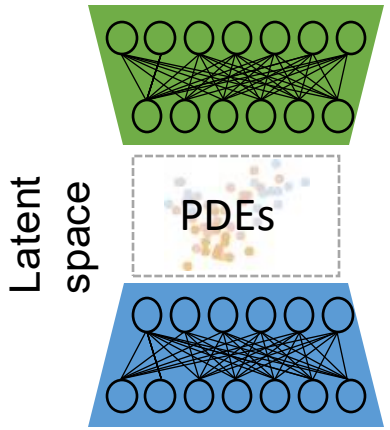
NSF-Simons Center for
Multiscale Cell Fate Research
UC Irvine

TIGON generating unmeasured single cells

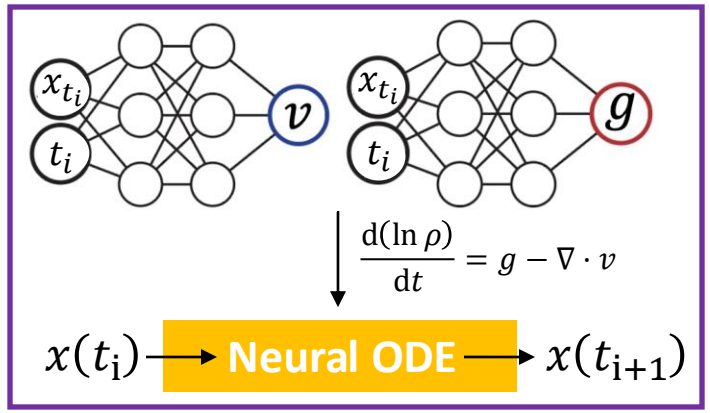
input



Self-supervised learning



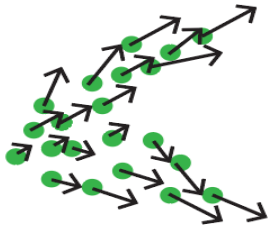
$$\partial_t \rho(x,t) + \nabla \cdot \big(v(x,t)\rho(x,t)\big) = g(x,t)\rho(x,t)$$



Output

Velocity/Trajectory

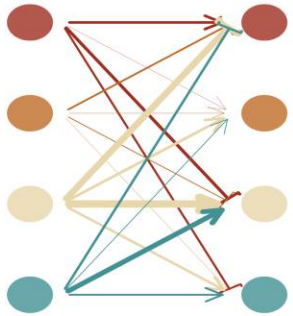
$v(x,t)$



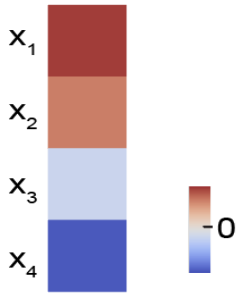
Growth

$g(x,t)$

Gene regulatory networks

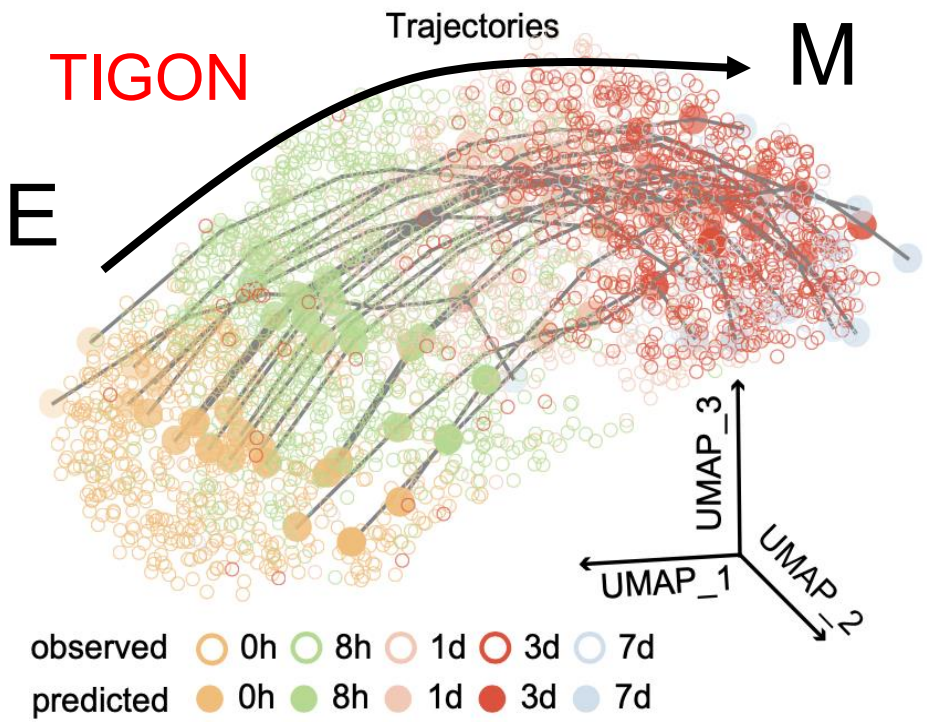


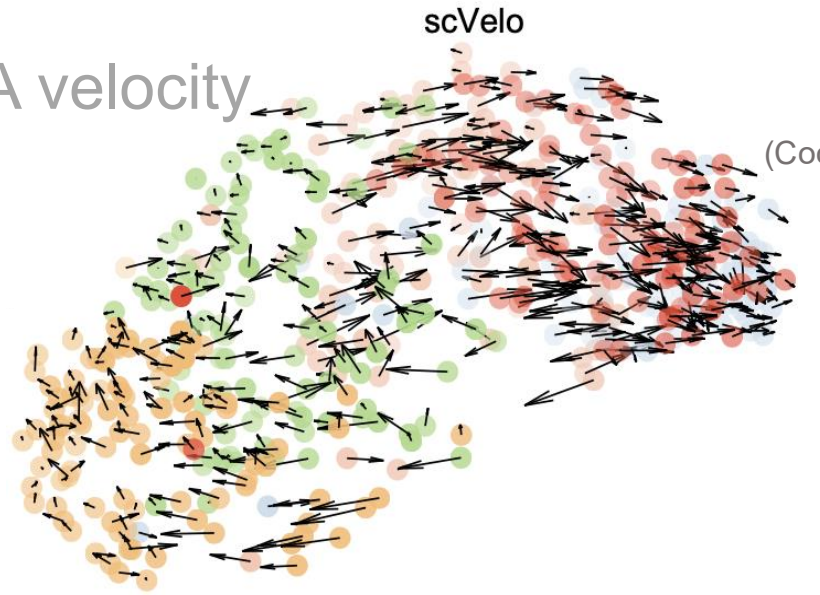
Growth-related genes



(**Sha** et al. Nature Machine Intelligence 2024)

TIGON on EMT data





RNA velocity

(Cook et al.,2020)

Inconsistent with the **real** time data

nature machine intelligence

2024

Article

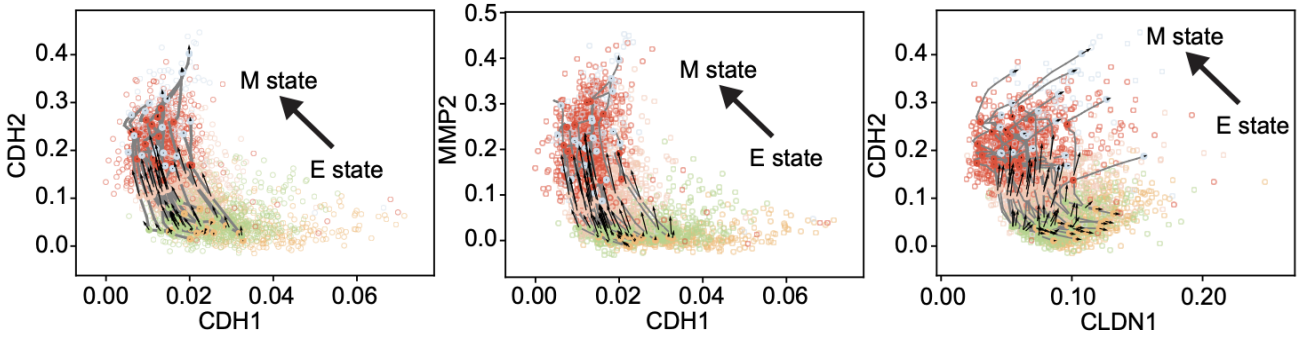
https://doi.org/10.1038/

Reconstructing growth and dynamic trajectories from single-cell transcriptomics data

Received: 8 February 2023

Accepted: 25 October 2023

Yutong Sha[1], Yuchi Qiu[2], Peijie Zhou[1] & Qing Nie[1,3,4]



Using LLM-type model to analyze single-cell data

words attention ~ cell-cell interaction

Using LLM methods to develop interactive analytics tools

Synthesizing multiple methods and making more user friendly

Using Generative AI in the lab

Ali Mortazavi

Dept to be: Systems Biology

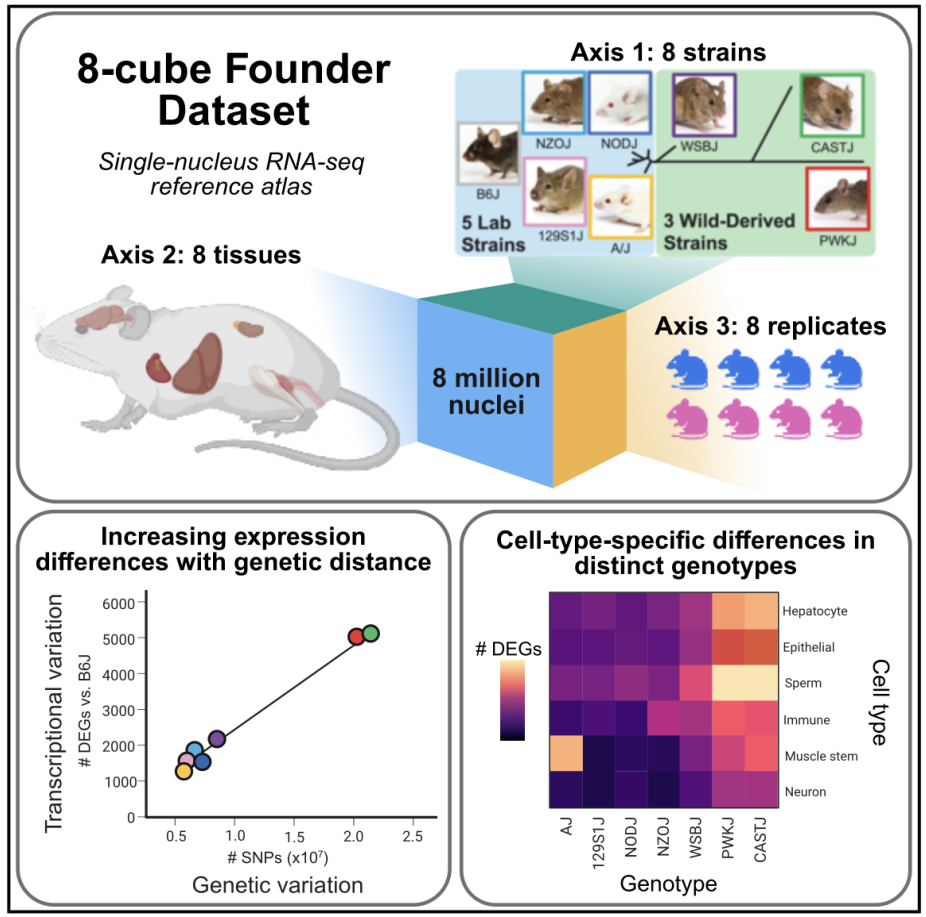
* most of the figures for this presentation were generated by Gemini Pro 3.1

My lab generates both large scale genomics datasets and software tools to use them

Cell Genomics

Resource

Systematic cell-type resolved transcriptomes of 8 tissues in 8 lab and wild-derived mouse strains capture global and local expression variation



(Rebboah et al, 2026)

Bioinformatics, 2018, 1–3
doi: 10.1093/bioinformatics/bty483
Advance Access Publication Date: 15 June 2018
Applications Note

Gene expression

TranscriptClean: variant-aware correction of indels, mismatches and splice junctions in long-read transcripts

Dana Wyman[1,2] **and Ali Mortazavi**[1,2,*]

[1]Department of Developmental and Cell Biology and [2]Center for Complex Biological Systems, UC Irvine, Irvine, CA, 92697, USA

Bioinformatics, 2020, 1–2
doi: 10.1093/bioinformatics/btaa836
vance Access Publication Date: 29 September 2020
Applications Note

OXFORD

Gene expression

Swan: a library for the analysis and visualization of long-read transcriptomes

Fairlie Reese [1,2] **and Ali Mortazavi** [1,2,*]

Bioinformatics, 2023, **39(7)**, btad415
https://doi.org/10.1093/bioinformatics/btad415
Advance access publication 3 July 2023
Applications Note

Gene expression

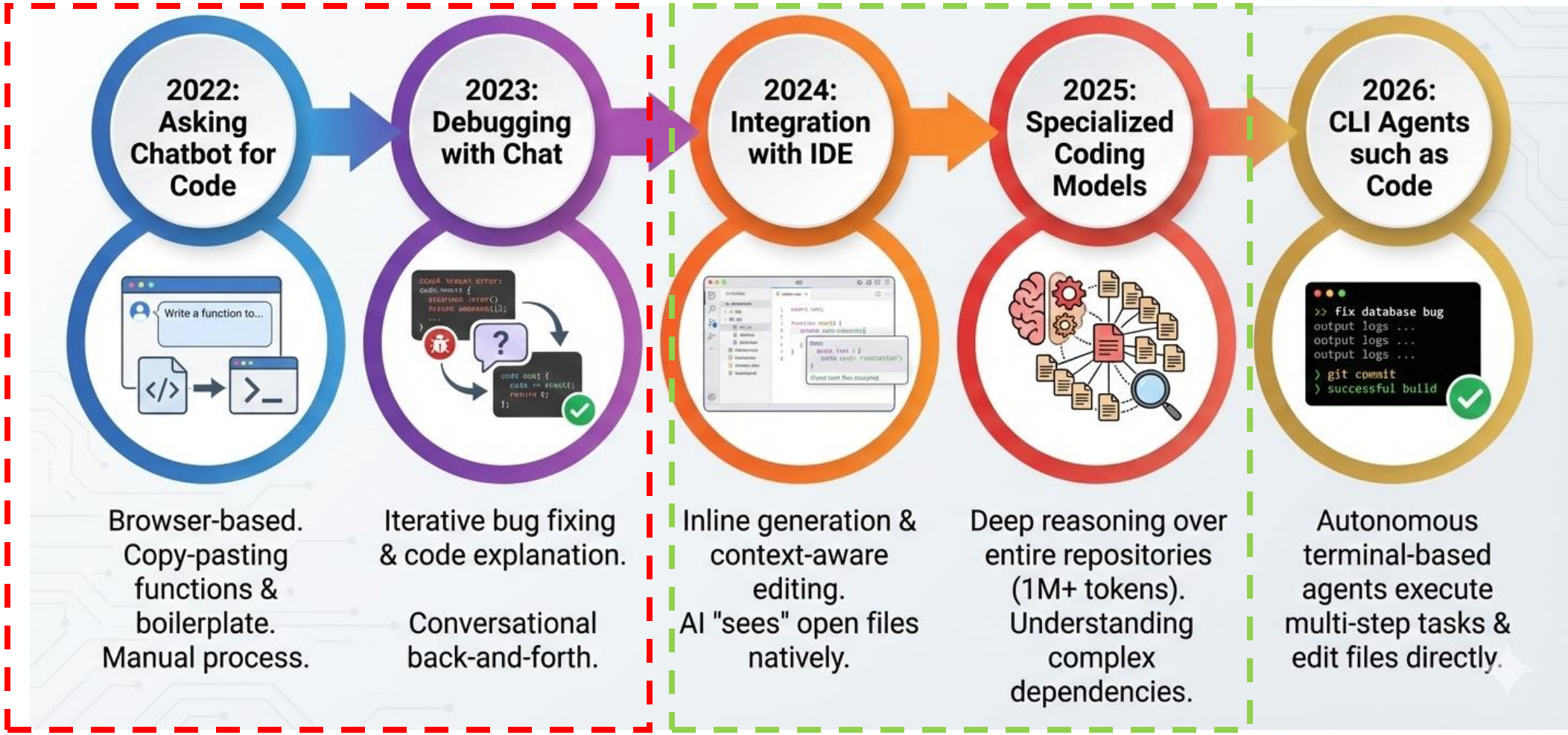
PyWGCNA: a Python package for weighted gene co-expression network analysis

Narges Rezaie[1,2], **Farlie Reese** [1,2], **Ali Mortazavi** [1,2,*]

[1]Department of Developmental and Cell Biology, UC Irvine, Irvine, CA 92697, United States
[2]Center for Complex Biological Systems, UC Irvine, Irvine, CA 92697, United States

Brave New World: all students in my lab are using AI as a coding partner

"zotGPT" + Large Models



Most of my students are here

One is here
(Me too)

Frontier Models are outperforming PhD-level scientists on benchmarks

Benchmark		Gemini 3.1 Pro Thinking (High)	Gemini 3 Pro Thinking (High)	Sonnet 4.6 Thinking (Max)	Opus 4.6 Thinking (Max)	GPT-5.2 Thinking (xhigh)	GPT-5.3-Codex Thinking (xhigh)
Humanity's Last Exam Academic reasoning (full set, text + MM)	No tools Search (blocklist) + Code	**44.4%** 51.4%	37.5% 45.8%	33.2% 49.0%	40.0% **53.1%**	34.5% 45.5%	— —
ARC-AGI-2 Abstract reasoning puzzles	ARC Prize Verified	**77.1%**	31.1%	58.3%	68.8%	52.9%	—
GPQA Diamond Scientific knowledge	No tools	**94.3%**	91.9%	89.9%	91.3%	92.4%	—
Terminal-Bench 2.0 Agentic terminal coding	Terminus-2 harness Other best self-reported harness	**68.5%** —	56.9% —	59.1% —	65.4% —	54.0% 62.2% (Codex)	64.7% **77.3%** (Codex)
SWE-Bench Verified Agentic coding	Single attempt	80.6%	76.2%	79.6%	**80.8%**	80.0%	—
SWE-Bench Pro (Public) Diverse agentic coding tasks	Single attempt	54.2%	43.3%	—	—	55.6%	**56.8%**
LiveCodeBench Pro Competitive coding problems from Codeforces, ICPC, and IOI	Elo	**2887**	2439	—	—	2393	—
SciCode Scientific research coding		**59%**	56%	47%	52%	52%	—
APEX-Agents Long horizon professional tasks		**33.5%**	18.4%	—	29.8%	23.0%	—
GDPval-AA Elo Expert tasks		1317	1195	**1633**	1606	1462	—
τ2-bench Agentic tool use	Retail Telecom	90.8% **99.3%**	85.3% 98.0%	91.7% 97.9%	**91.9%** **99.3%**	82.0% 98.7%	— —
MCP Atlas Multi-step workflows using MCP		**69.2%**	54.1%	61.3%	59.5%	60.6%	—
BrowseComp Agentic search	Search + Python + Browse	**85.9%**	59.2%	74.7%	84.0%	65.8%	—
MMMU Pro Multimodal understanding and reasoning	No tools	80.5%	**81.0%**	74.5%	73.9%	79.5%	—
MMMLU Multilingual Q&A		**92.6%**	91.8%	89.3%	91.1%	89.6%	—
MRCR v2 (8-needle) Long context performance	128k (average) 1M (pointwise)	**84.9%** 26.3%	77.0% 26.3%	**84.9%** Not supported	84.0% Not supported	83.8% Not supported	— —

Methodology: deepmind.google/models/evals-methodology/gemini-3-1-pro

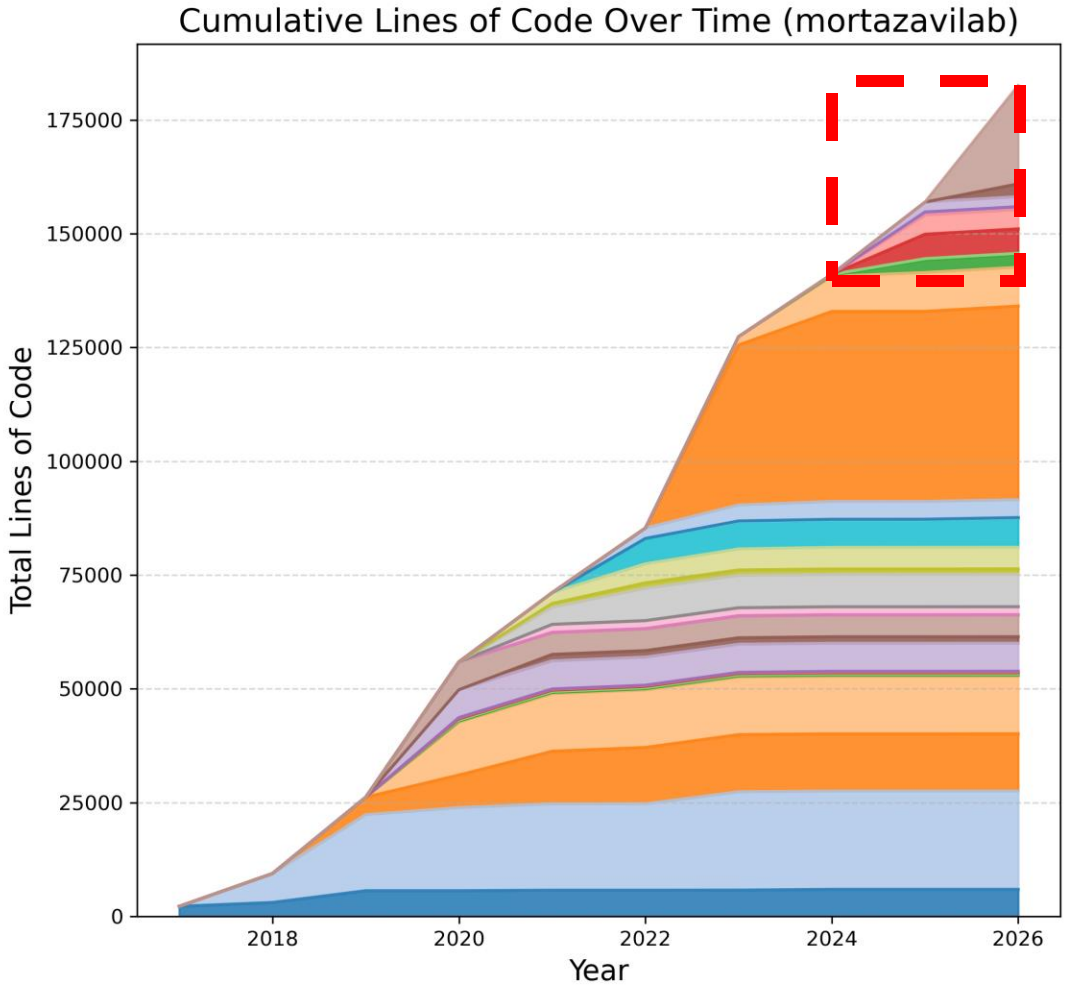
GPQA Diamond:

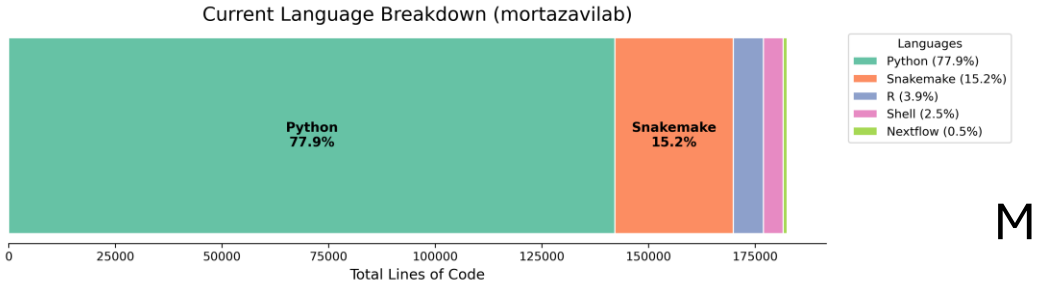
- a highly challenging subset of 198 PhD-level, "Google-proof" questions in biology, chemistry, and physics.
- the most rigorous, expert-curated questions designed to test advanced AI reasoning
- domain experts achieve approximately 65–81% accuracy
- non-experts struggle to perform significantly better than random guessing.

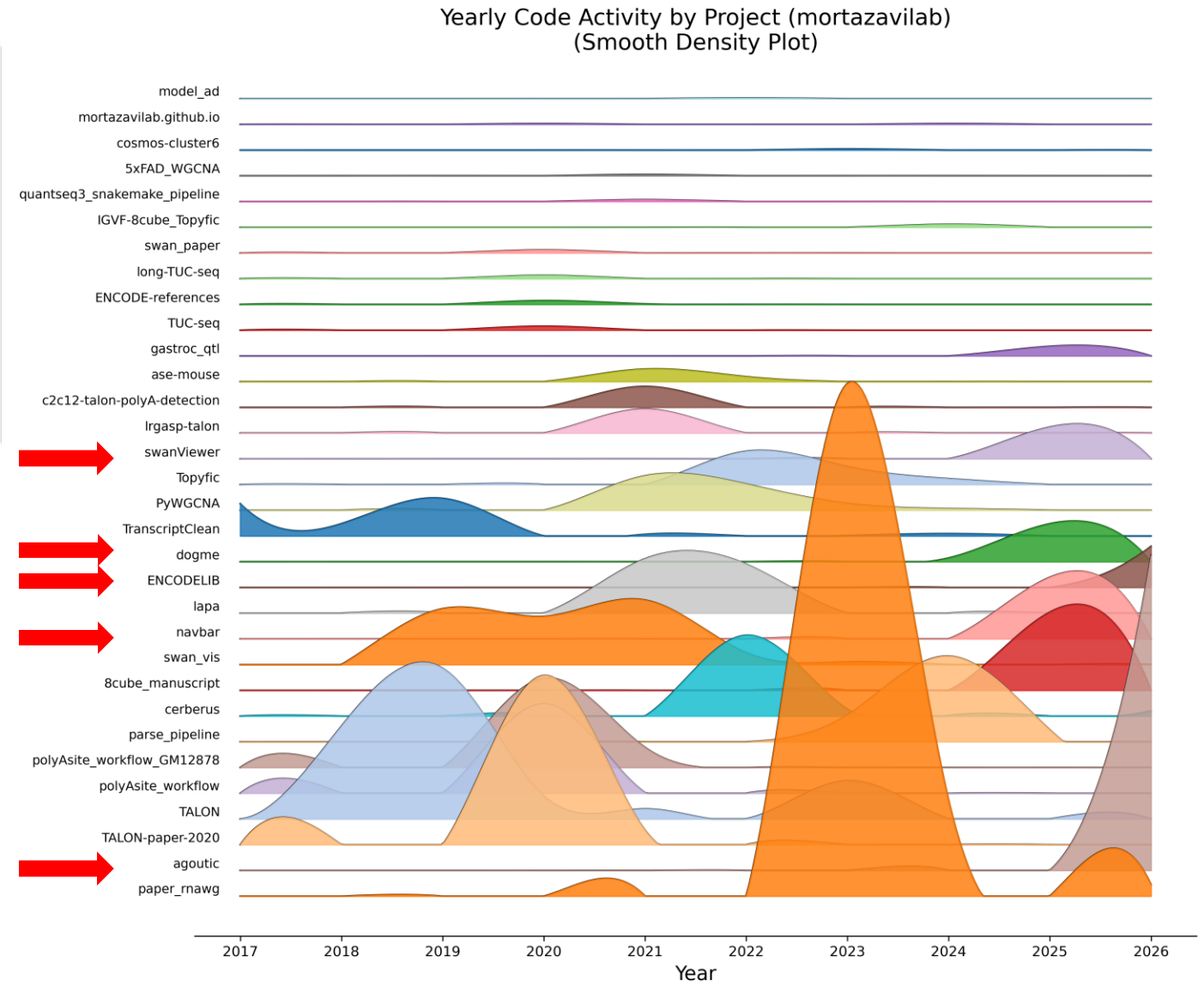
SciCode:

- scientist-curated benchmark designed to evaluate Large Language Models (LLMs) on realistic scientific research coding
- decomposes 80 complex, real-world scientific problems into 338 subproblems
- 10% are biological (modeling)
- Performance 2 years ago was 5%

The bulk of the 50k code deposited by my lab since 2024 was AI-assisted

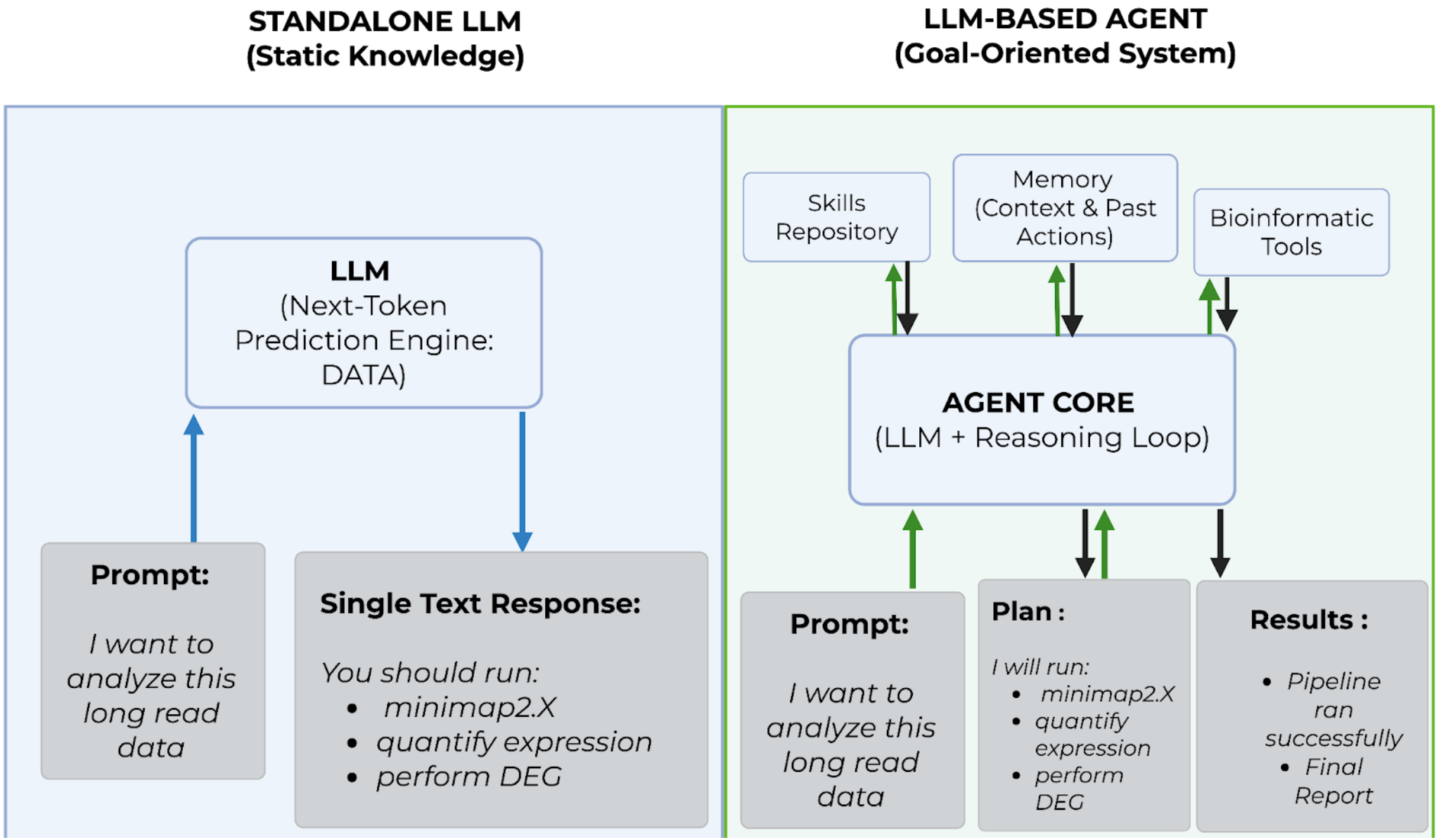






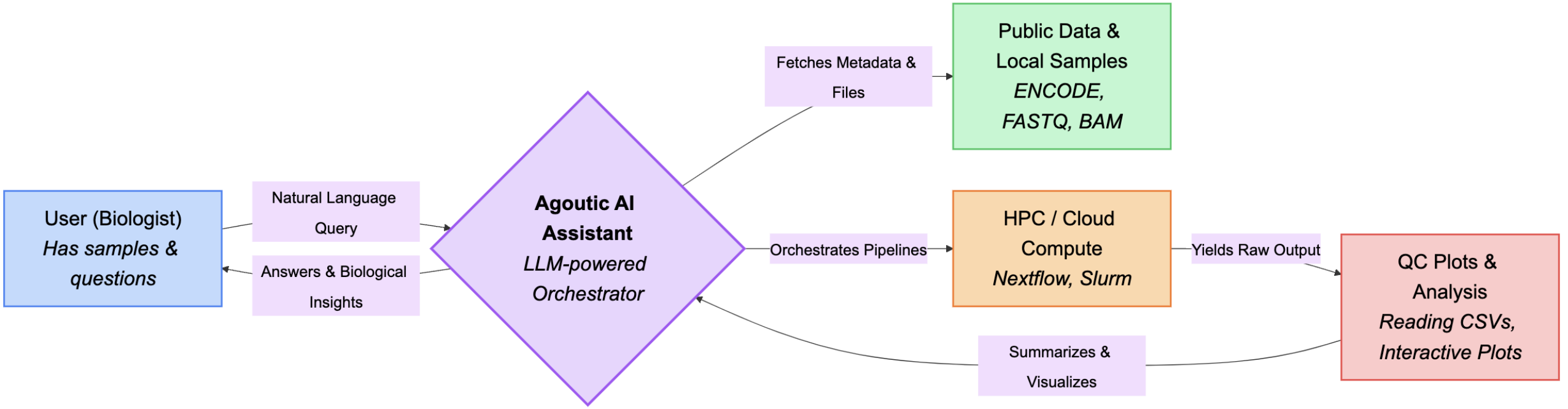
Most AI projects in 2025 were ~2-4k lines of code (GUIs, pipeline)

Moving from chatbots to autonomous agents



Building an agent to automate most bioinformatics analyses

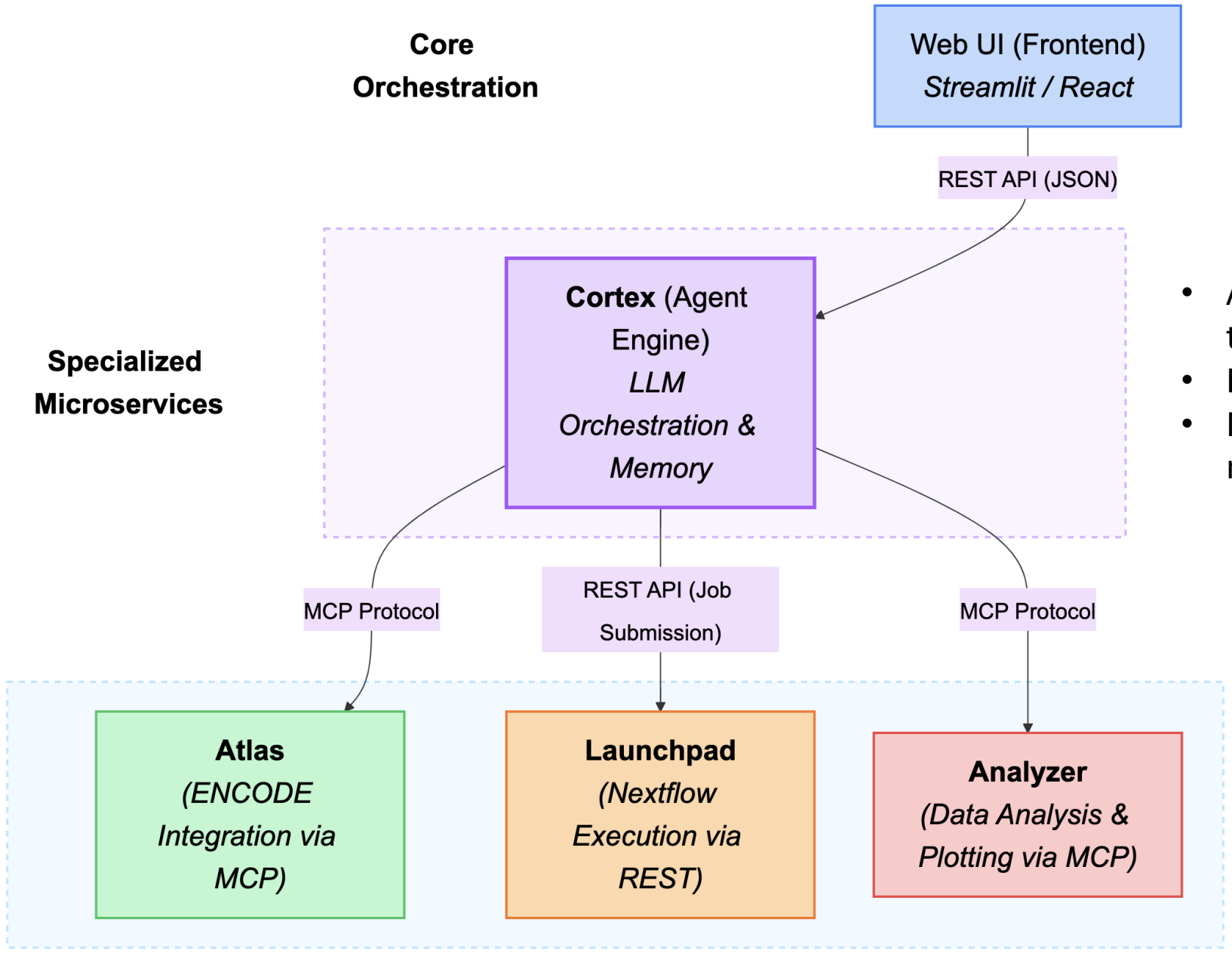
Being built first to process long-read data through our Dogme workflow,
but will eventually handle all our lab data



- Runs on a local LLM on NVIDIA GPUs
- Devstral-2 for now, can swap models

Started January 2026 after a couple of weeks of writing a product requirement document:
Currently 22k lines of code – growing ~11k / month – expect public release around end April

The Agoutic distributed architecture

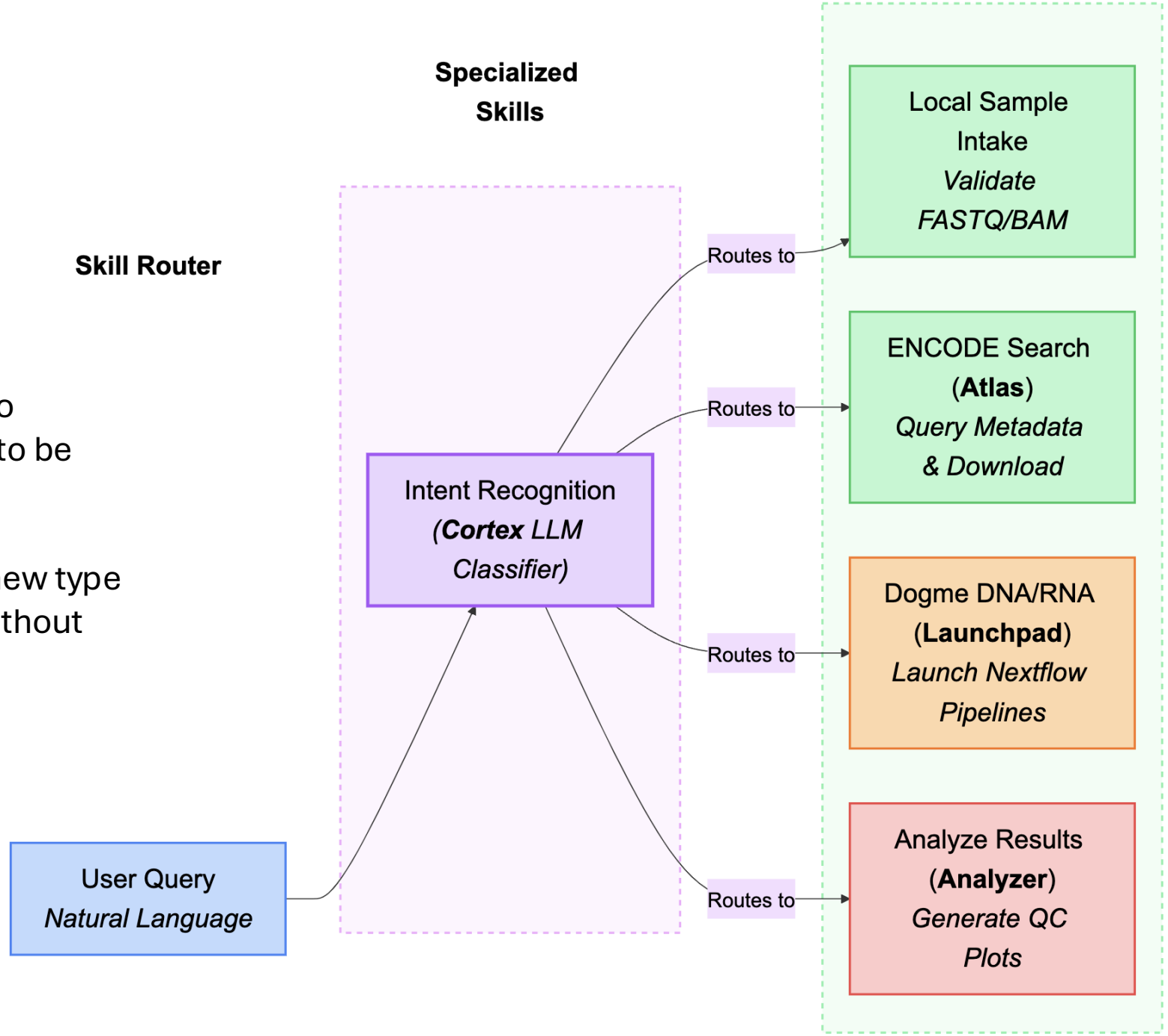


- All the individual parts are in place and tested
- Running the first test samples end-to-end
- Needs to be upgraded to run jobs remotely on HPC

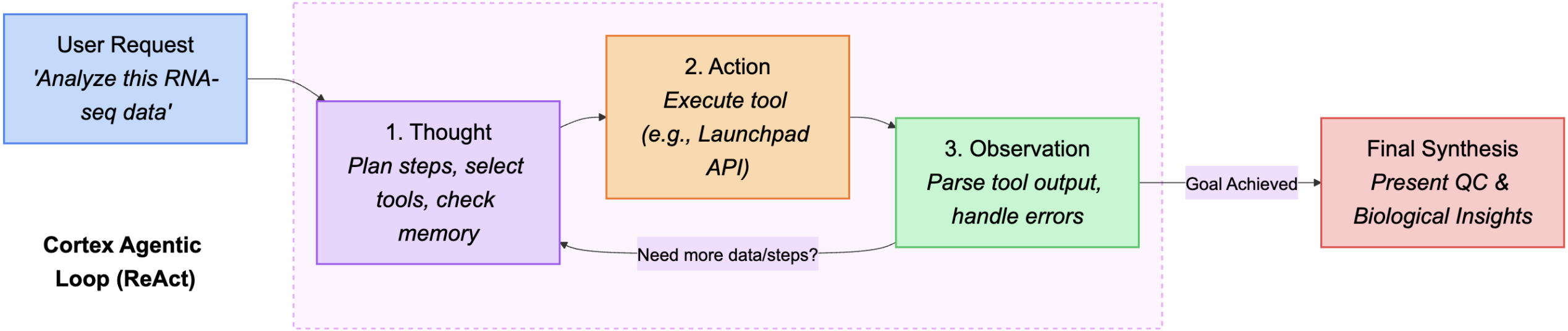
The skills routing system

Skills are detailed text files that allow us to specify how we want different data types to be processed and analyzed.

If we want to add a new type of data or a new type of analysis, we can add a new skills file without changing the underlying code.

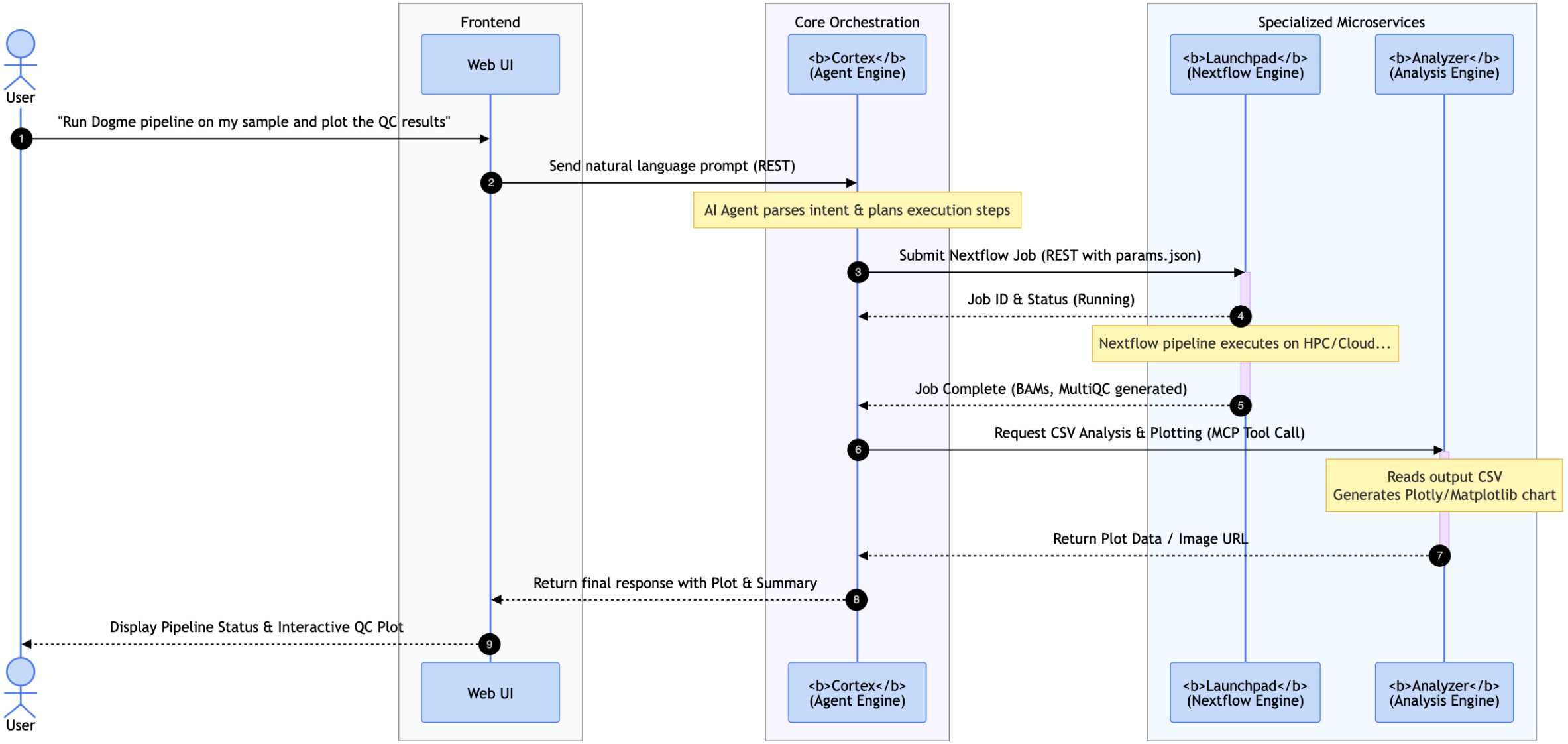


How Agoutic Breaks Down Complex Biological Queries into Actionable Steps



Goal is for someone not familiar with tools or bioinformatics to unlock the results in their data

Seamless Integration of Job Submission, Monitoring, and Visualization

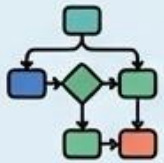


Managing Scientific Data Projects in 2026: A Modern Blueprint

From Raw Data to Reproducible Discoveries

1. AI-Assisted Pipeline Automation

- **Adopt CLI Agents:** Use terminal-based AI agents (like Claude Code or Aider) to quickly generate boilerplate scripts, troubleshoot massive log files, and refactor legacy lab code.



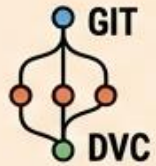
- **Standardize Workflows:** Transition loose scripts into workflow managers like Nextflow or Snakemake. This allows seamless scaling from a local laptop to an HPC cluster or the cloud.

2. Bulletproof Reproducibility

- **Containerize Everything:** Never rely on local machine setups (which lead to "it works on my machine" issues). Enforce the use of Docker or Singularity containers for all analysis steps.

- **Lock Environments:** Automate the generation of environment.yml or requirements.txt files so anyone can recreate the exact software state.

3. Strict Data Versioning & Provenance



- **Track Data Like Code:** Use DVC (Data Version Control) alongside Git. Git tracks the scripts; DVC tracks the massive datasets without bloating your repository.

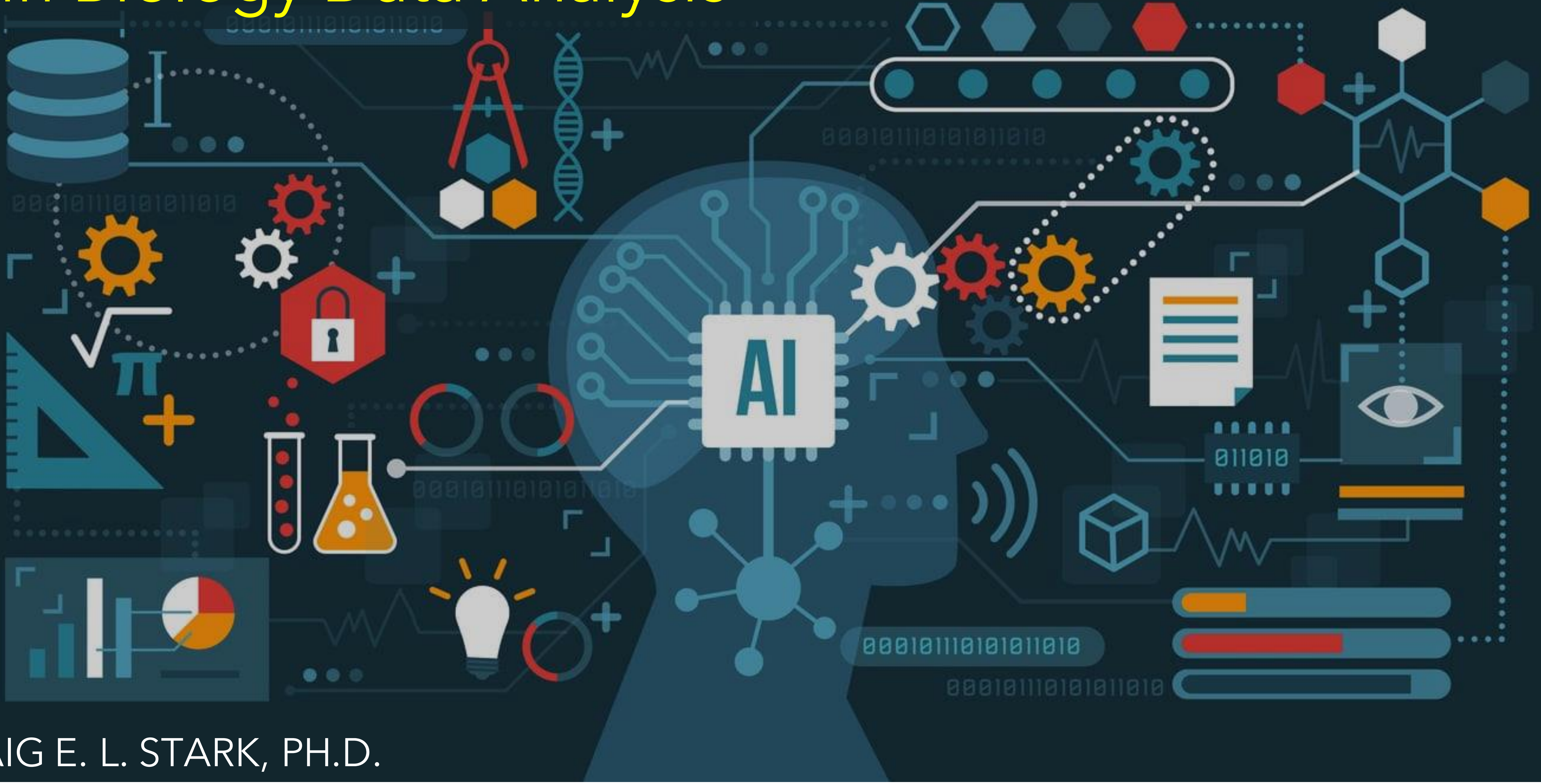
- **Immutable Raw Data:** Treat raw data as strictly read-only. Log the exact AI prompts and scripts used to create any derived datasets.

4. Shift to "Human-in-the-Loop" Review

- **Review, Don't Just Write:** As AI handles more of the heavy lifting in coding, shift your team's focus toward rigorous code review and verifying biological/scientific logic.

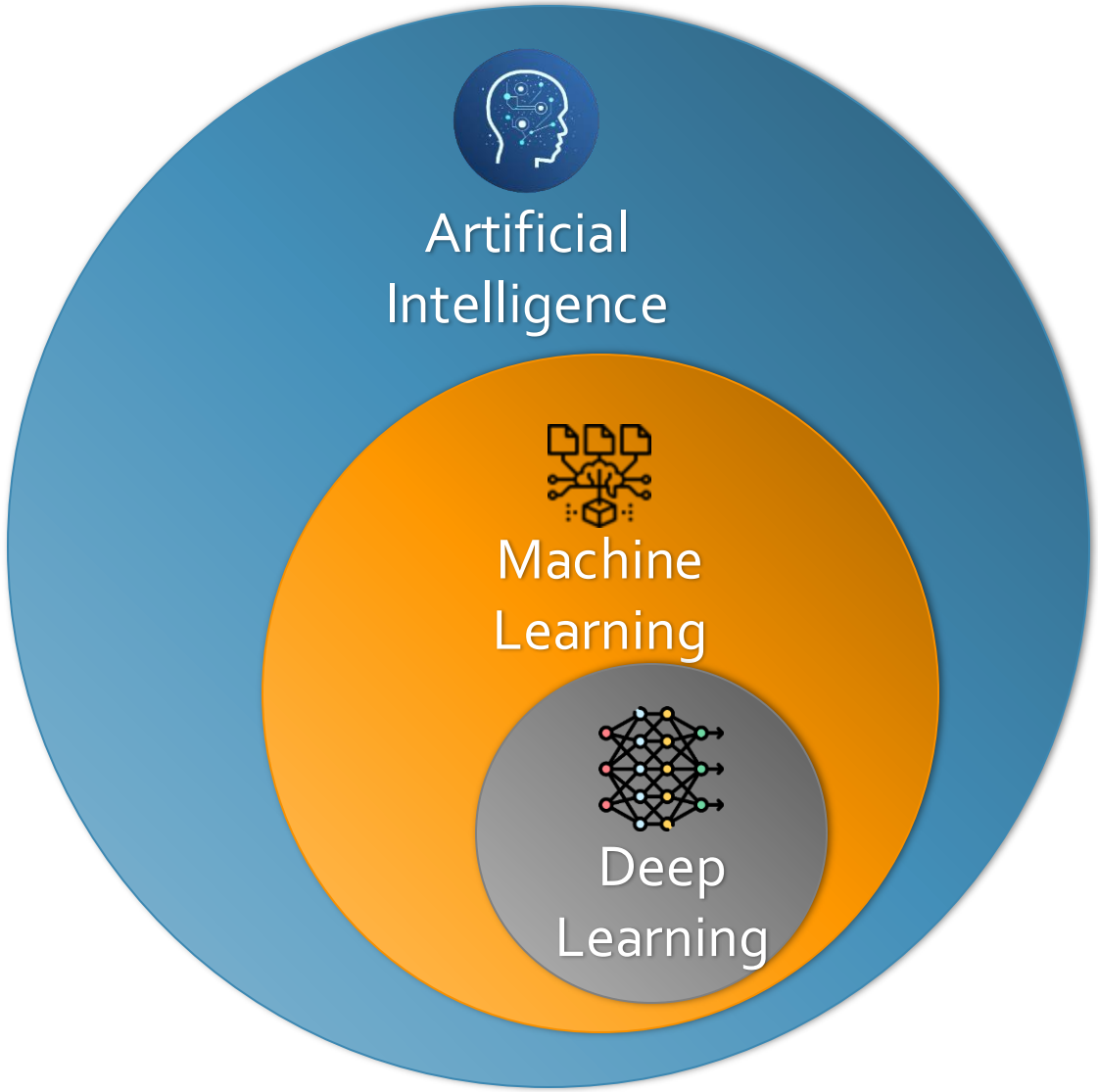
- **Lab Pull Requests:** Institute mandatory peer reviews for analysis pipelines within the lab before any paper is submitted.

AI in Biology Data Analysis



CRAIG E. L. STARK, PH.D.
PROFESSOR OF NEUROBIOLOGY AND BEHAVIOR
CHARLIE DUNLOP SCHOOL OF BIOLOGICAL SCIENCES
UNIVERSITY OF CALIFORNIA IRVINE

AI vs Machine Learning vs. Deep Learning



Artificial Intelligence (AI)
Systems designed to simulate human thinking and decision making

Machine Learning (ML)
Systems that learn to perform a task from vast amounts of data rather than direct instruction

Deep Learning
Systems that learn from vast amounts of data via multi-layer neural networks

Can we diagnose dementia from low-burden data?

Received: 9 June 2023 | Revised: 19 September 2023 | Accepted: 26 September 2023

DOI: 10.1002/dad2.12494

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Improving clinical efficiency in screening for cognitive impairment due to Alzheimer's

Yueqi Ren[1,2] | Babak Shahbaba[1,3] | Craig E. L. Stark[1,4]

2: Medium	**(12)**	vegetable and animal naming, digit span
	MRI (155)	Gross volumes, regional volumes, and cortical thicknesses
3: High	**Genetic testing (1)**	ApoE allele carrier status
	Clinical Dementia Rating (8)	CDR global, CDR sum of boxes, subdomain scores

STARK LAB

Can we diagnose dementia from low-burden data?

Received: 9 June 2023 | Revised: 19 September 2023 | Accepted: 26 September 2023
DOI: 10.1002/dad2.12494
RESEARCH ARTICLE
Alzheimer's & Dementia® Diagnosis, Assessment & Disease Monitoring

Improving clinical efficiency in screening for cognitive impairment due to Alzheimer's

Yueqi Ren[1,2] | Babak Shahbaba[1,3] | Craig E. L. Stark[1,4]

Tier	Modality (N)	Features
1: Low	Demographics (10)	Age, sex, education level
		Race, ethnicity, language(s) spoken
		Marital status, living situation
	Patient history (71)	Tobacco use, alcohol use, medications
		Cardiovascular conditions and comorbidities (e.g., congestive heart failure, stroke, diabetes)
		Family history of dementia
		Physical exam (e.g., heart rate, blood pressure, BMI)
	Behavioral surveys (39)	NACC Functional Assessment Scale
		Neuropsychiatric Inventory Questionnaire
		Geriatric Depression Scale
	Neuropsychological testing (1)	Mini-Mental State Exam
2: Medium	Neuropsychological testing (12)	Logical Memory II – Delayed, Trails A and B, Boston Naming Test, vegetable and animal naming, digit span
	MRI (155)	Gross volumes, regional volumes, and cortical thicknesses
3: High	Genetic testing (1)	ApoE allele carrier status
	Clinical Dementia Rating (8)	CDR global, CDR sum of boxes, subdomain scores

STARK LAB

Can we diagnose dementia from low-burden data?

Received: 9 June 2023 | Revised: 19 September 2023 | Accepted: 26 September 2023
DOI: 10.1002/dad2.12494

RESEARCH ARTICLE

Alzheimer's & Dementia® Diagnosis, Assessment & Disease Monitoring

Improving clinical efficiency in screening for cognitive impairment due to Alzheimer's

Yueqi Ren[1,2] | Babak Shahbaba[1,3] | Craig E. L. Stark[1,4]

Data

Preprocessing

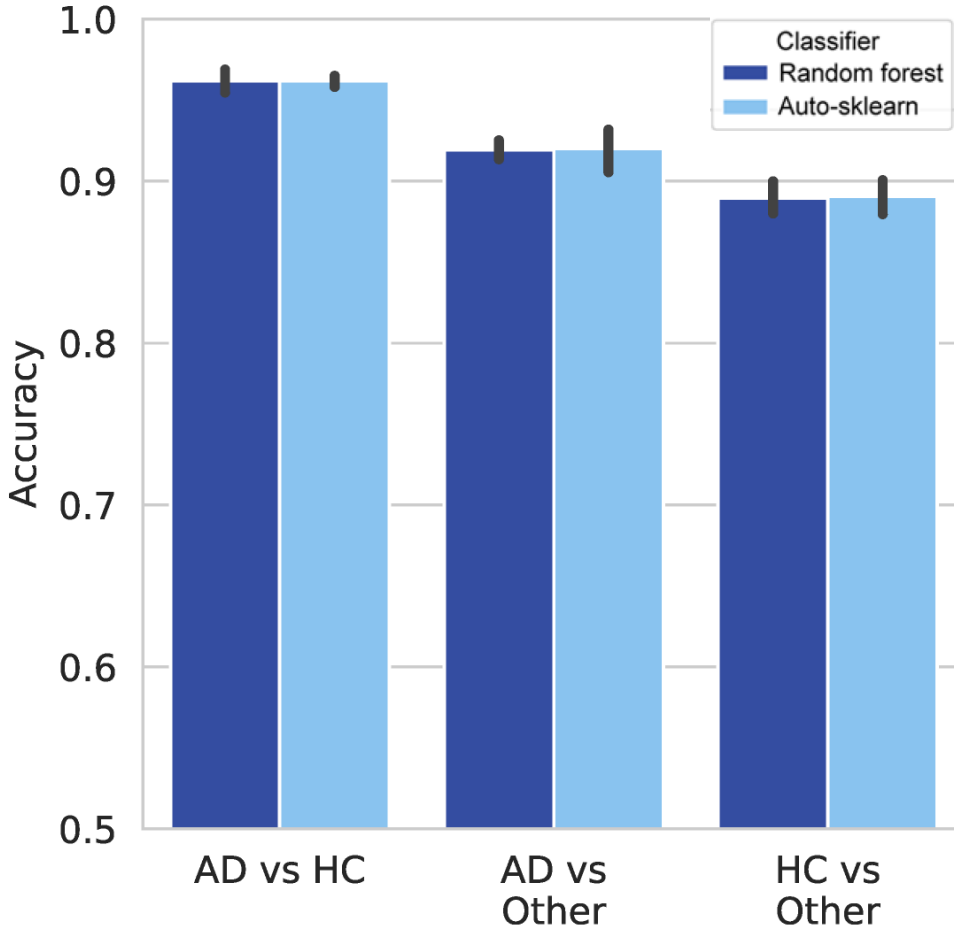
5-fold nested CV

SMOTE oversampling to correct class imbalance

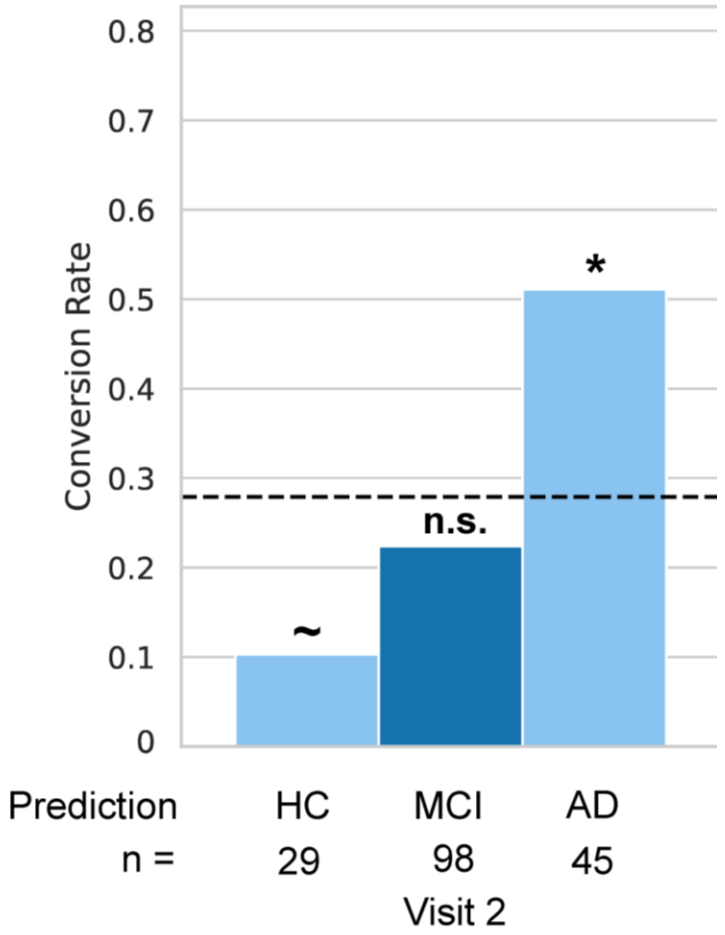
Feature selection using extra trees

Machine learning classifier

2-Way Accuracy for Tier 1



"Errors" predict future progression



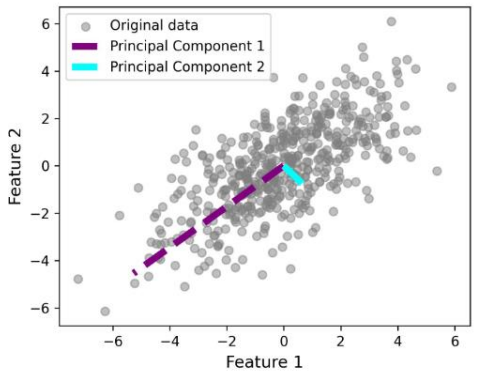
STARK LAB

frontiers | Frontiers in Dementia

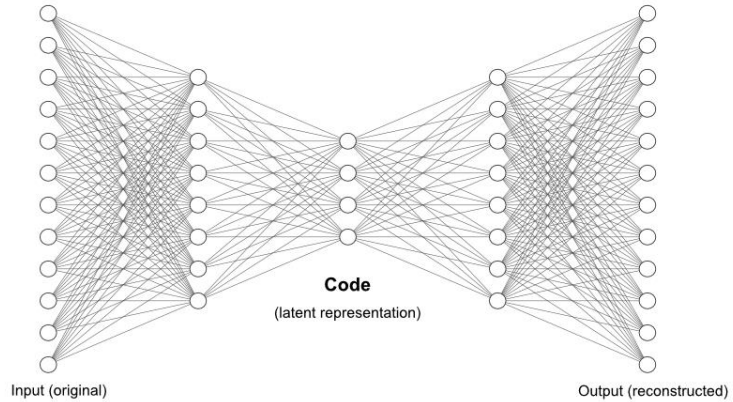
Alzheimer's disease detection using data fusion with a deep supervised encoder

Minh Trinh[1†], Ryan Shahbaba[2†], Craig Stark[3,4] and Yueqi Ren[4,5*]

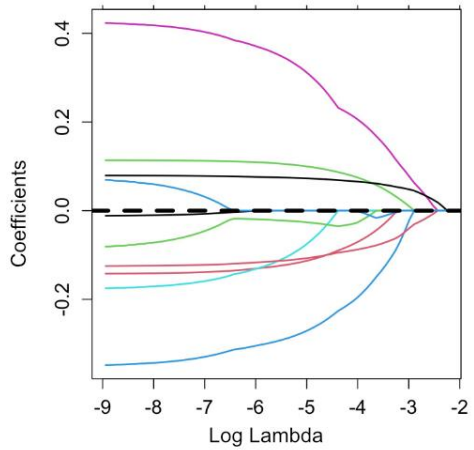
a) PCA



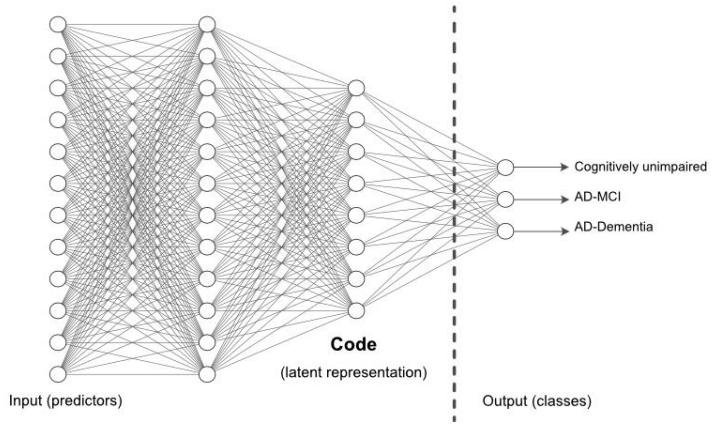
b) Autoencoder (AE)

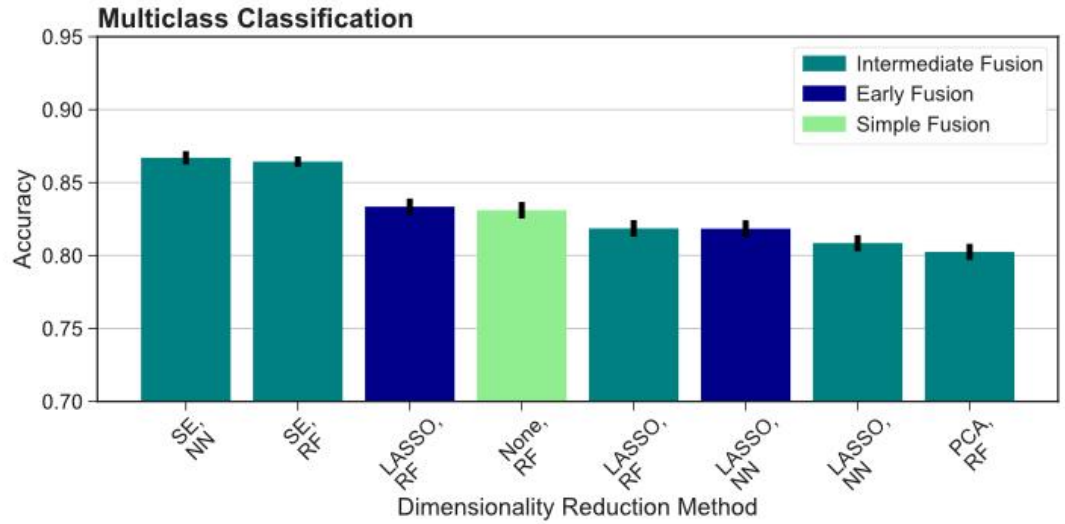


c) LASSO



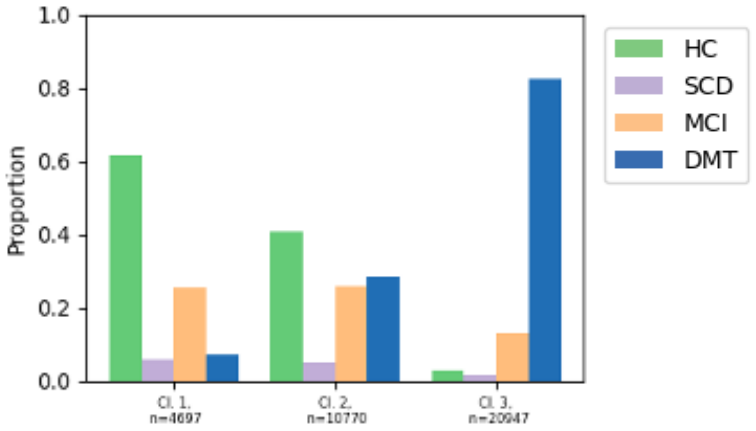
d) Supervised Encoder (SE)





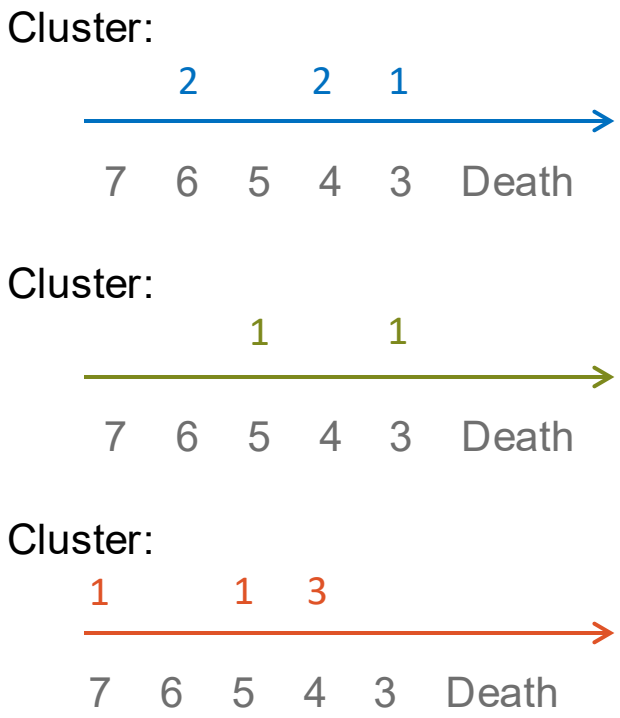
Can we identify meaningful clinical subgroups that reflect neuropathology?

Using data from all subjects without autopsy, **generate representative clusters** of the **disease landscape**:

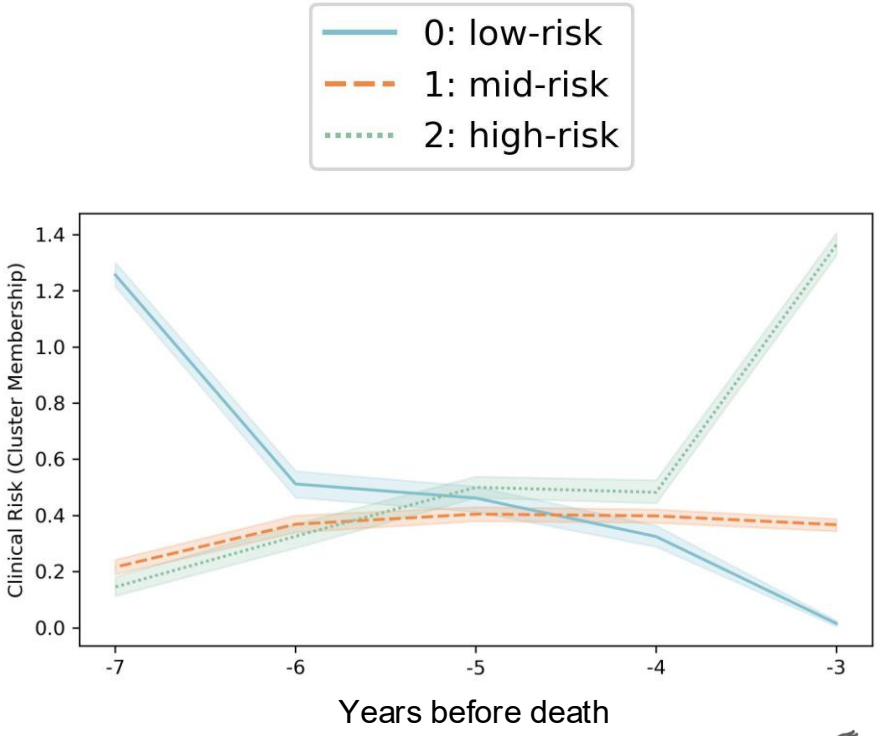


Mostly healthy ⟷ Mostly impaired

Assign subjects with autopsy to clusters within the existing landscape **at each time point**:



Track **trajectory of cluster assignments** for subjects with autopsy to find **3 risk groups**:

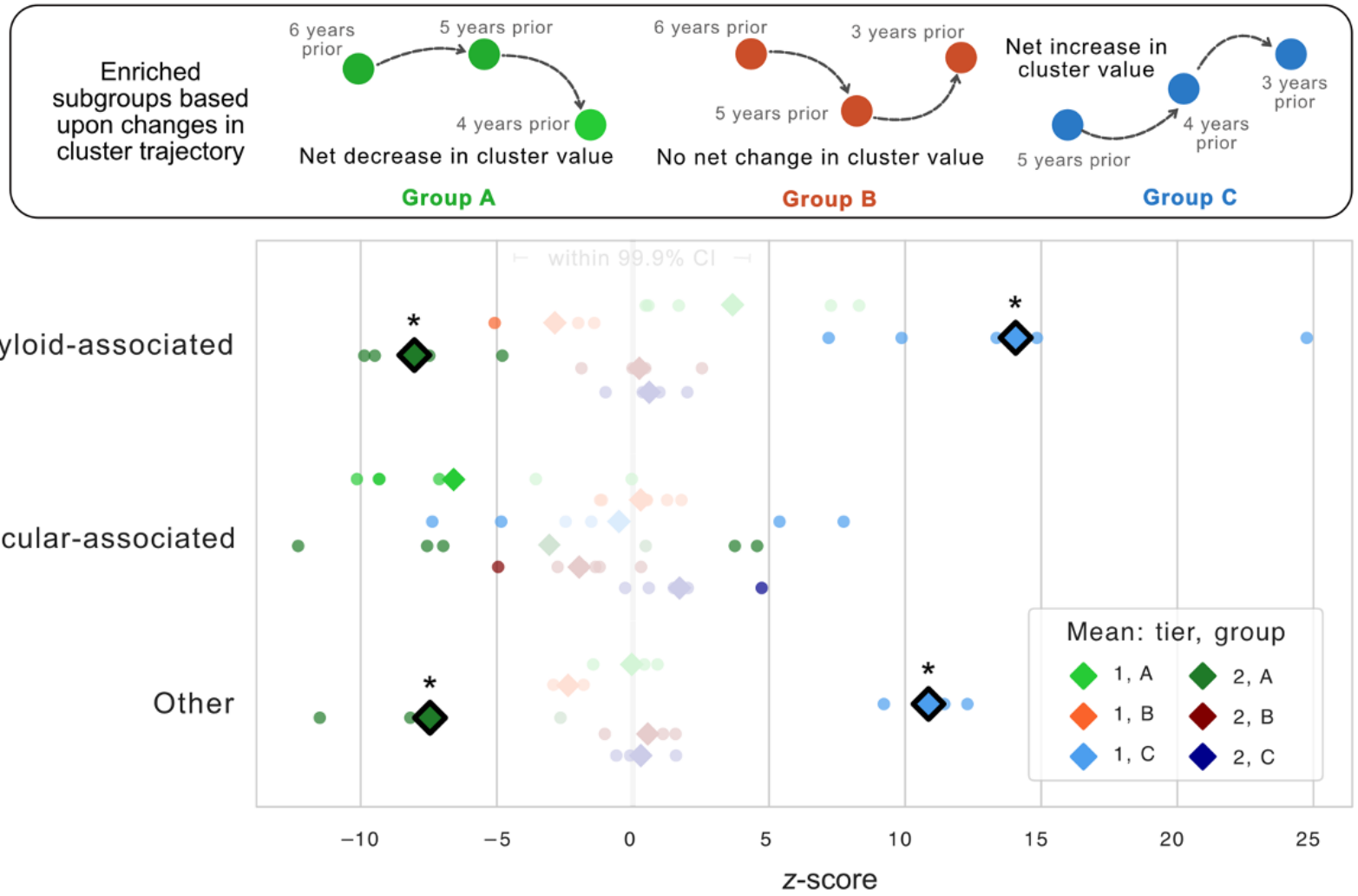


Can we predict pathology?

Received: 5 March 2025 | Revised: 30 May 2025 | Accepted: 7 July 2025
DOI: 10.1002/alz.70539
Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION
RESEARCH ARTICLE

Identifying dementia neuropathology using low-burden clinical data

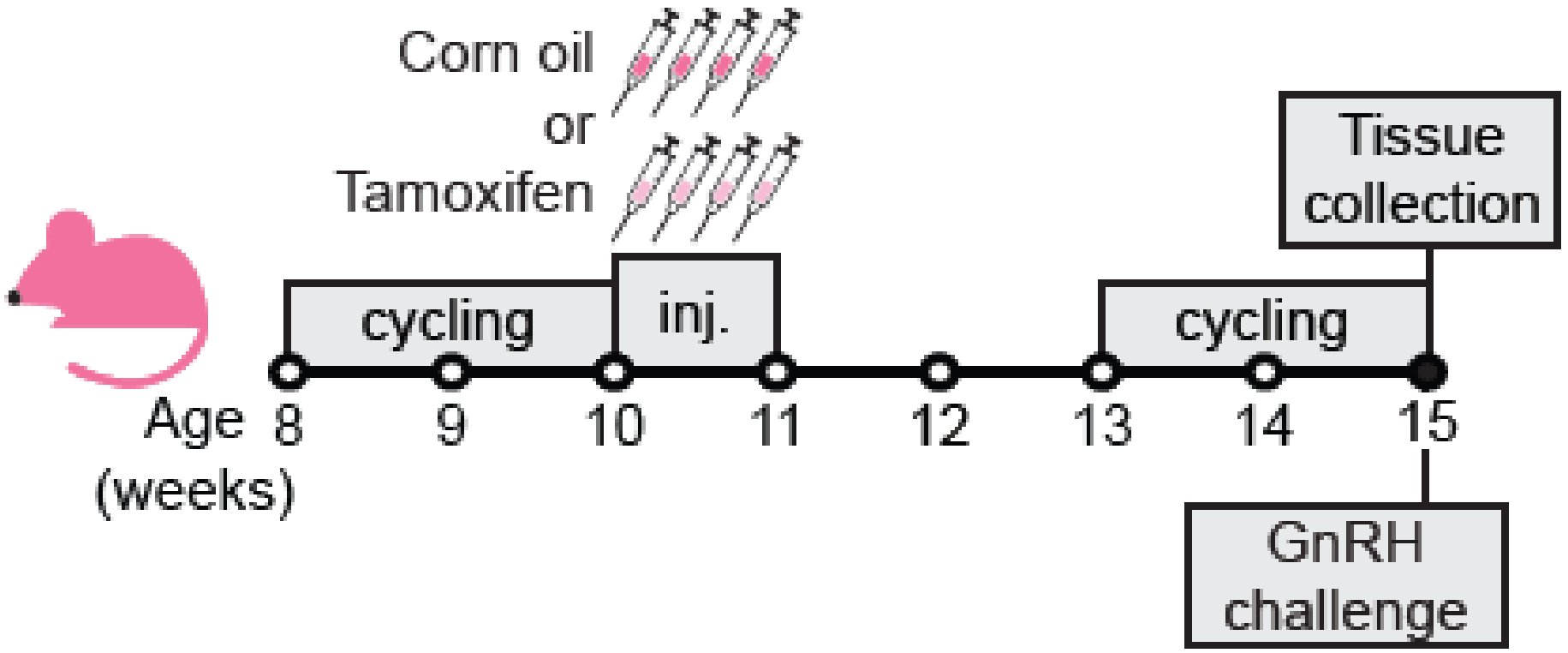
Yueqi Ren[1] | Babak Shahbaba[2] | Craig E. L. Stark[3]



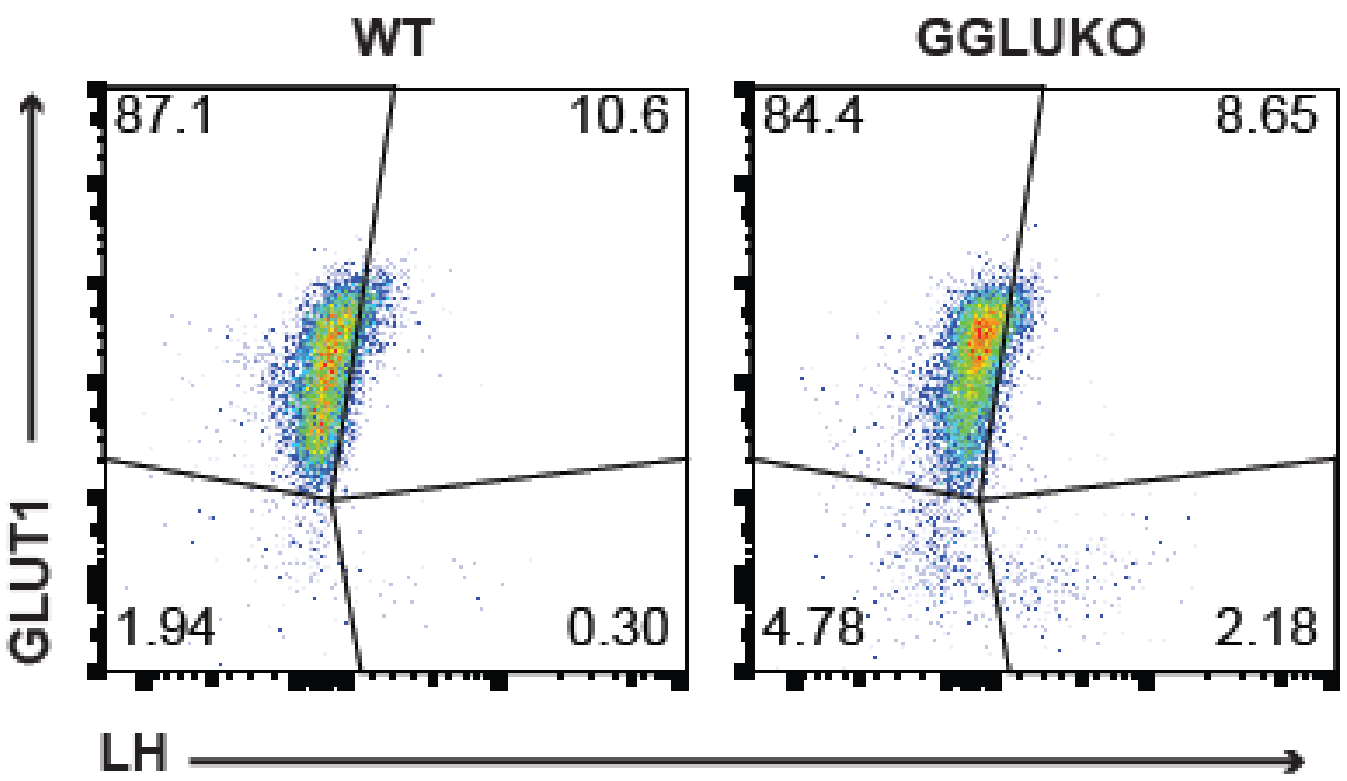
Meaningful data from a suboptimal experiment

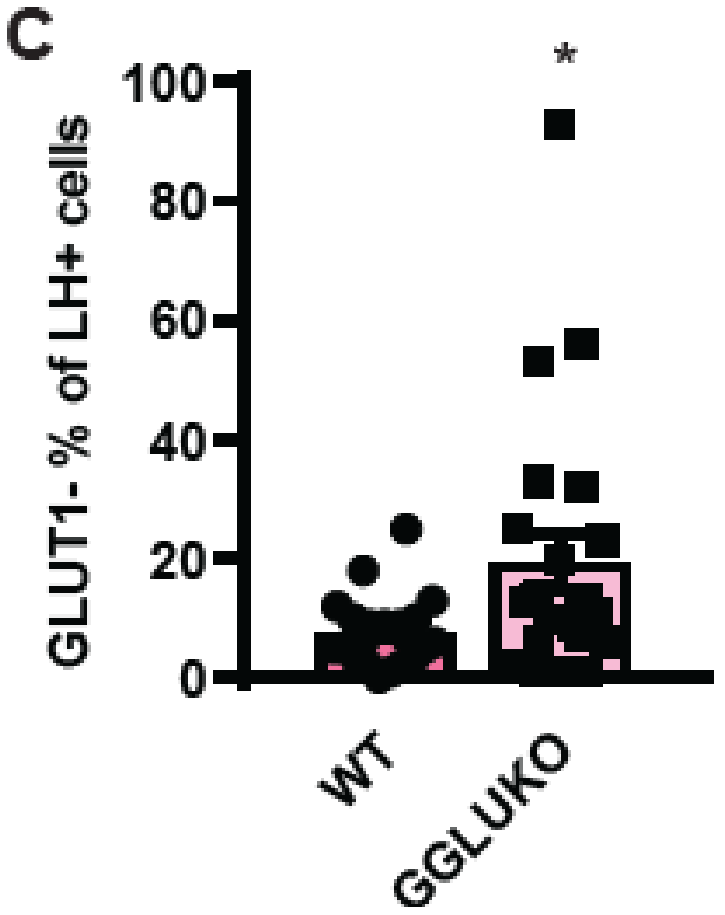
Dequina Nicholas
MB&B

Conditional KO of GLUT1



KO Efficiency Sucked!!





Help ZotGPT!?!?

1. Looking at LH response in KO

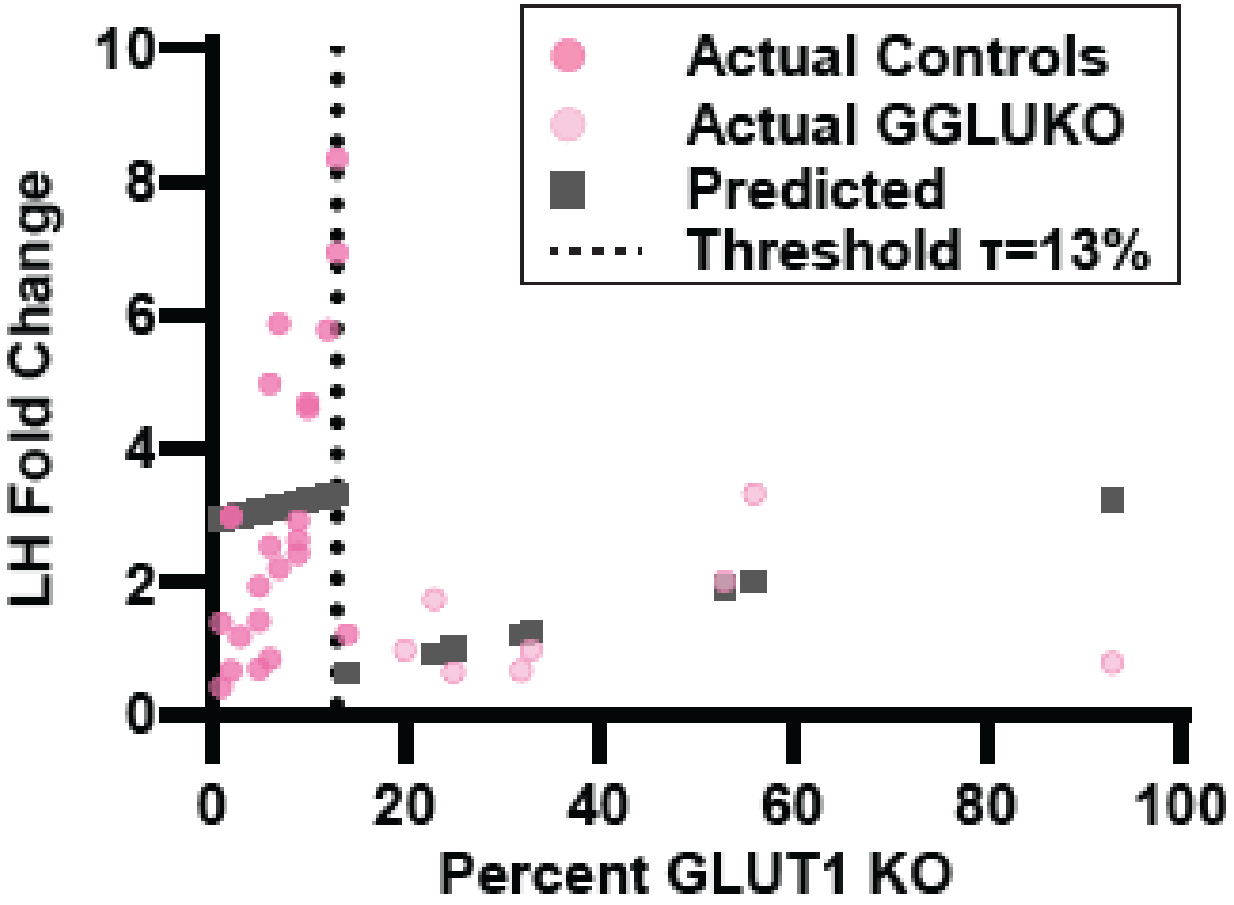
2. If I make an arbitrary cut-off for %KO, then the loss of response is significant

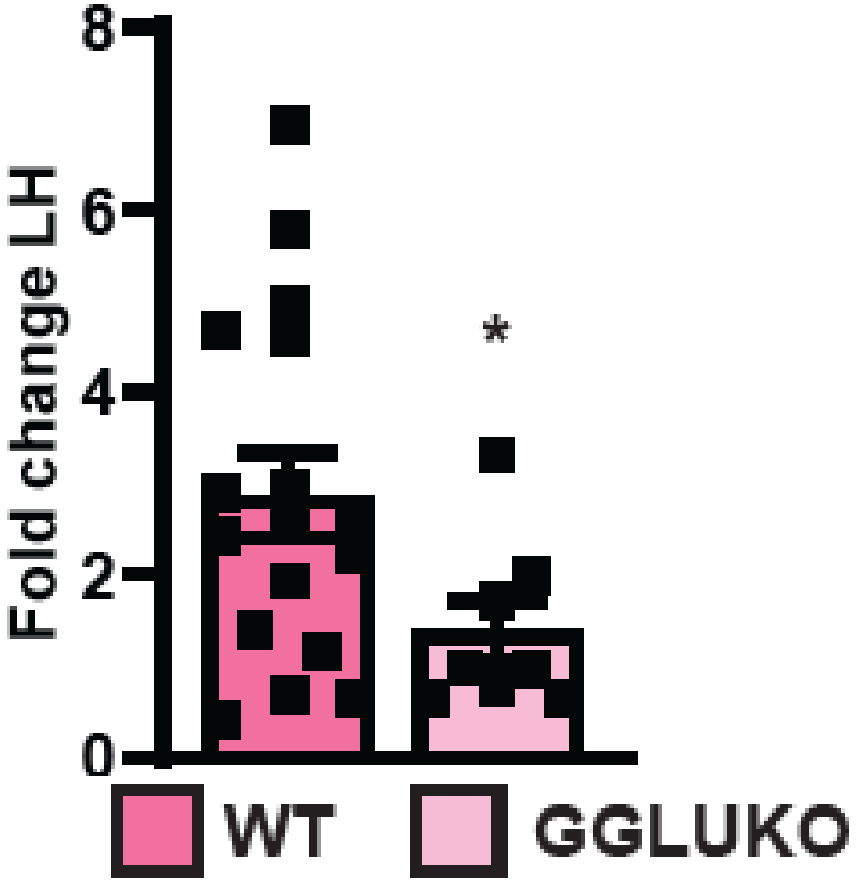
3. I can see patterns in the data, but don't know how to analyze

Threshold Analysis!!
Quantitative, specialized type of sensitivity analysis used to identify the specific value of an input parameter at which a decision, outcome, or system behavior changes significantly

Used in economics and healthcare

Threshold Analysis

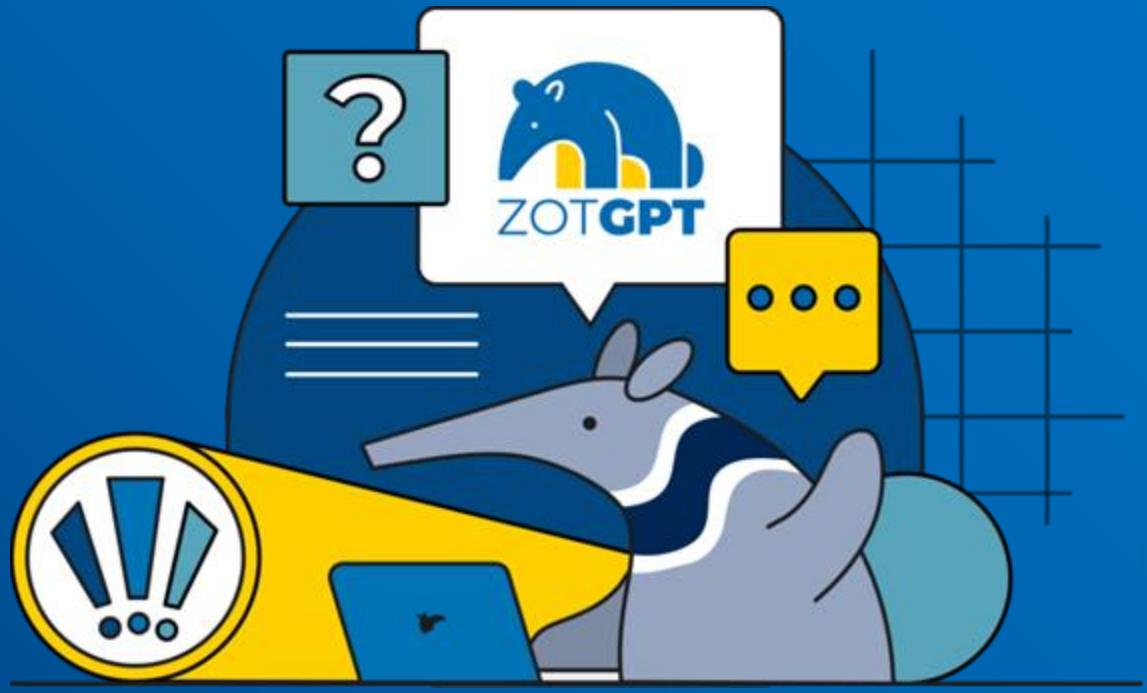




Slide deck: bit.ly/uci-ai-research

ZotGPT

Generative AI Tools & Resources at UCI



UC Irvine

Slide deck: bit.ly/uci-ai-research

ZotGPT

AI by UCI, for UCI

- **Secure:** Secure, UCI-governed environment for protected data (P3)
- **Equitable:** Available to all faculty, staff, and students at no cost
- **Cutting-edge:** Access to the latest and most advanced AI models
- **UCI-Tailored:** Tools and training customized for UC Irvine's academic, research, and administrative needs

UC Irvine

Slide deck: bit.ly/uci-ai-research

ZotGPT Chat

Powered by Industry Leading AI Models

- Claude Opus 4.6
- GPT-5.2
- Gemini 2.5 Pro
- DeepSeek R1
- Mistral Large 3

And many more!

UC Irvine

Slide deck: bit.ly/uci-ai-research

Categories of Generative AI Use Cases


Ideation


Composition


Roleplay


Organization


Summarization


Interpretation


Visualization


Coding


Evaluation


Translation


Retrieval


Personalization


Coaching


Prototyping


Storytelling

UC Irvine

Slide deck: bit.ly/zotgpt26

ZotGPT in Action

Content Summarizer

Virtual Ideator

Data Interpreter

Data Visualizer

ZotGPT Assistant

UC Irvine

What is Creator?

ZotGPT's solution for custom AI assistants

- No-code custom chatbot platform for faculty and staff
- Develop chatbots trained on your data
- Secure environment for private use or large deployments at scale

Slide deck: bit.ly/uci-ai-research

ZotGPT Academy

Free Gen AI Training for UC Irvine Faculty + Staff

- Foundational skills for working with Generative AI
- Features video tutorials and e-learning modules
- Four modules currently available:
  - **Module 1:** Generative AI Fundamentals
  - **Module 2:** Prompt Engineering
  - **Module 3:** Augmenting your Work with Generative AI
  - **Module 4:** Creating Custom AI Assistants

UC Irvine

Slide deck: bit.ly/uci-ai-research

Thank you!

Questions?

Dominic Slauson
Generative AI Specialist

My website
codaptivelabs.com

Contact
slausond@uci.edu

LinkedIn
dominic-slauson

ZotGPT website
zotgpt.uci.edu

Contact
oit-ai@uci.edu

UCI Beall
Applied Innovation

Measurably improving lives through access to UCI-developed knowledge

UC Irvine

An AI That's A Better Friend

Bill Tomlinson's AI doesn't just chat. It remembers, it cares, and it might just help us save ourselves.

Read

Writing and AI Literacy for 50 Million Students

UC Irvine researchers envisioned using AI, alongside education, cultural studies, and visual arts, to provide elementary school students personalized instruction, engage them in the process of writing, and build AI literacy through interactive storytelling.

UCI Startup

StoryAI

PoP AWARD $50K

UC Irvine

Utilized by 25 Million Students

Developed from ground-breaking UC Irvine research, ALEKS is a web-based, artificially intelligent assessment and learning system that, in 1997, marked a significant leap in personalized education at scale.

To date, ALEKS has been utilized by over 25 million students.

UCI Startup

ALEKS

(McGraw-Hill Education)

UC Irvine

Learn More

Reassurance for 400K Preterm Infants

Feeding is a critical milestone for premature babies, requiring coordination of muscles, sensory responses, and breathing. Without objective tools, doctors rely on observation, often prolonging NICU stays. A UC Irvine NICU physical therapist developed a tool providing real-time data on feeding readiness, helping families bring their babies home sooner, safely.

UCI Startup

NeoCare Innovations

PoP AWARD $80K

Additional Backing

$256K STTR (NSF)
$944K SBIR (NSF)
$320K Other Grants

Discover More

UC Irvine

Our Future to Design

UC Irvine's NarrA.I.tive Story Studio and the Ethical Frontiers of Artificial Intelligence

Read

NarrAItive

UC Irvine

Story Studio

- A hybrid lab that combines traditional film studio elements with emerging AI technologies.
- Provides incubation for new technologies, experiential learning opportunities, and campus storytelling services.

Why Science Needs a Story

UC Irvine
Beall Applied Innovation

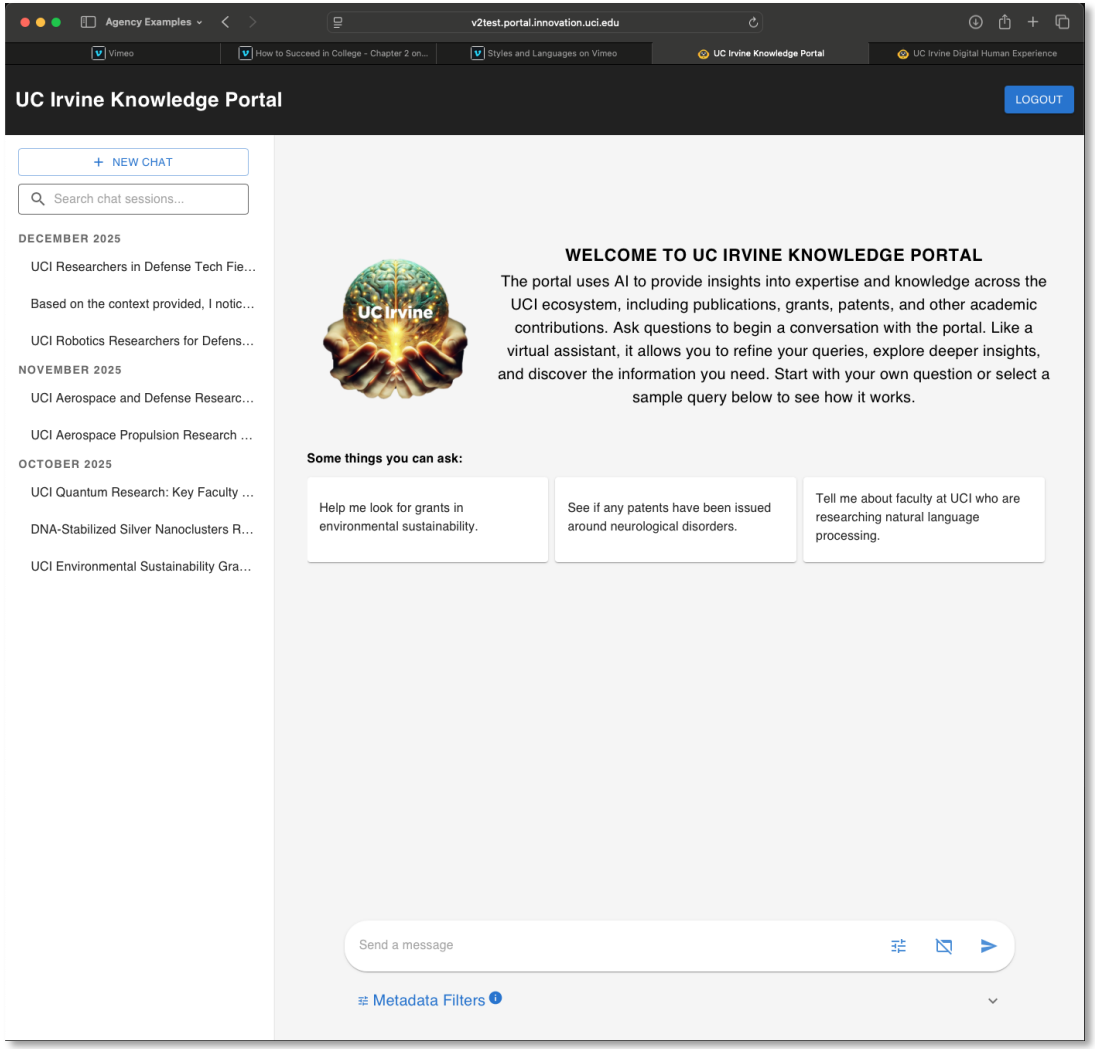
NarrA.I.tive

UC Irvine Portal

Portal

UC Irvine
Beall Applied Innovation

The Knowledge Portal is a RAG environment that combines publications, grants, patents and other academic contributions from UC Irvine.



UC Irvine
Beall Applied Innovation

- Request: Identify faculty with experience in vision, cognition, and athletic performance—specifically, whether enhancements in visual input or visual clarity can produce measurable improvements in physical reaction time.

- Oakley signed research agreement within 1 week.

Academic Unit	Faculty Member	Expertise / Research Focus	Why They're a Good Fit for Oakley's Project	Contact Information*
Cognitive Sciences (School of Social Sciences)	Charles E. Wright, PhD	Human motor control, aimed movements, visuomotor coordination, response latency, psychophysical task design	Connects visual perception to motor output; ideal for quantifying and modeling physical reaction times under visual conditions	cewright@uci.edu • Profile
Cognitive Sciences (School of Social Sciences)	Alyssa A. Brewer, MD PhD	Visual neuroscience, sensory mapping, cortical representation of perception, attention	Can design and validate visual stimuli or lens-based manipulations; bridges optical and cortical effects of vision improvements	aabrewer@uci.edu • Profile
Cognitive Sciences (School of Social Sciences)	Megan Peters, PhD	Perception, metacognition, computational neuroscience, psychophysics, decision-making	Expert in perceptual decision processes; can help link improved visual input to measurable behavioral speed or accuracy	megan.peters@uci.edu • Lab Site
Cognitive Sciences (School of Social Sciences)	Michael D. Lee, PhD	Computational cognitive modeling, evidence-accumulation models of decision and reaction time, Bayesian inference	Models RT data statistically; ideal for quantifying how visual clarity or cue salience affects speed/accuracy trade-offs	mdlee@uci.edu • Profile
Biomedical			Provides	