LLMs for Graduate Students

Jordan K, Lotte B, Maddie B, Nate G, Remi S

and the rest of the LLM ‘Reading’ Group
(Andrea P, Andrew J, Artyom L, Aryaman B, Dominic G, Kate M, Marcelo S, Matt M, Melvyn N)

What is happening?

LLMs are transforming how physics is done.

Experiment

Principles and Techniques for Agentic Research in Experimental Physics [2610.17193]
"The use of artificial intelligence (AI) in physics research is increasingly common, yet few guidelines exist for its effective use. [...]"

Theory

TheoPhysEval: Assessing AI agents at autonomous theoretical physics problem solving [2610.16733]
"[...] agents can accurately perform concrete calculations required across theoretical physics research."

https://arxiv.org/abs/2610.16733
https://arxiv.org/abs/2610.17193

So what do we do about it?

We must hold regular meetings between grads and faculty to discuss and adapt.

- Empower ourselves to make LLMs useful
- Encourage responsible uses in research
- Manage the effects on learning and teaching
- Ensure equitable access to new technologies

"We have to think about keeping the field alive in the future: How are we training the next group of physicists? […] We have to balance the resources that we're using, which includes the cost of doing the AI investigations. […] Human physics is not over."

- Sarah Demers, Yale Physics Chair

https://www.quantamagazine.org/in-an-age-of-ai-a-physicist-seeks-what-endures-20260903/

Overview

01

Introduction

Slide 2-4



02

AI in Experiments

Slide 5-10



03

AI in Theory, Data Analysis, & Programming

Slide 11-13





04

Research

Slide 14-17



05

Learning & Teaching

Slide 18-21



06

Procurement

Slide 22

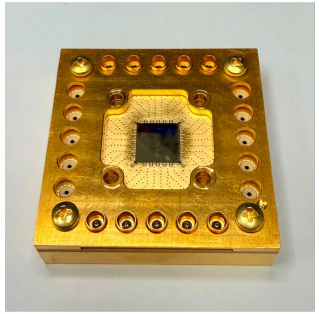


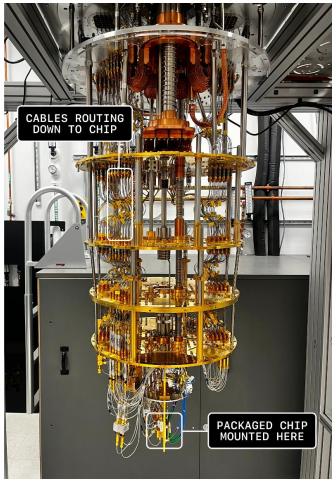
AI in Experiments

Nate

Agentic AI in Tabletop Experiments

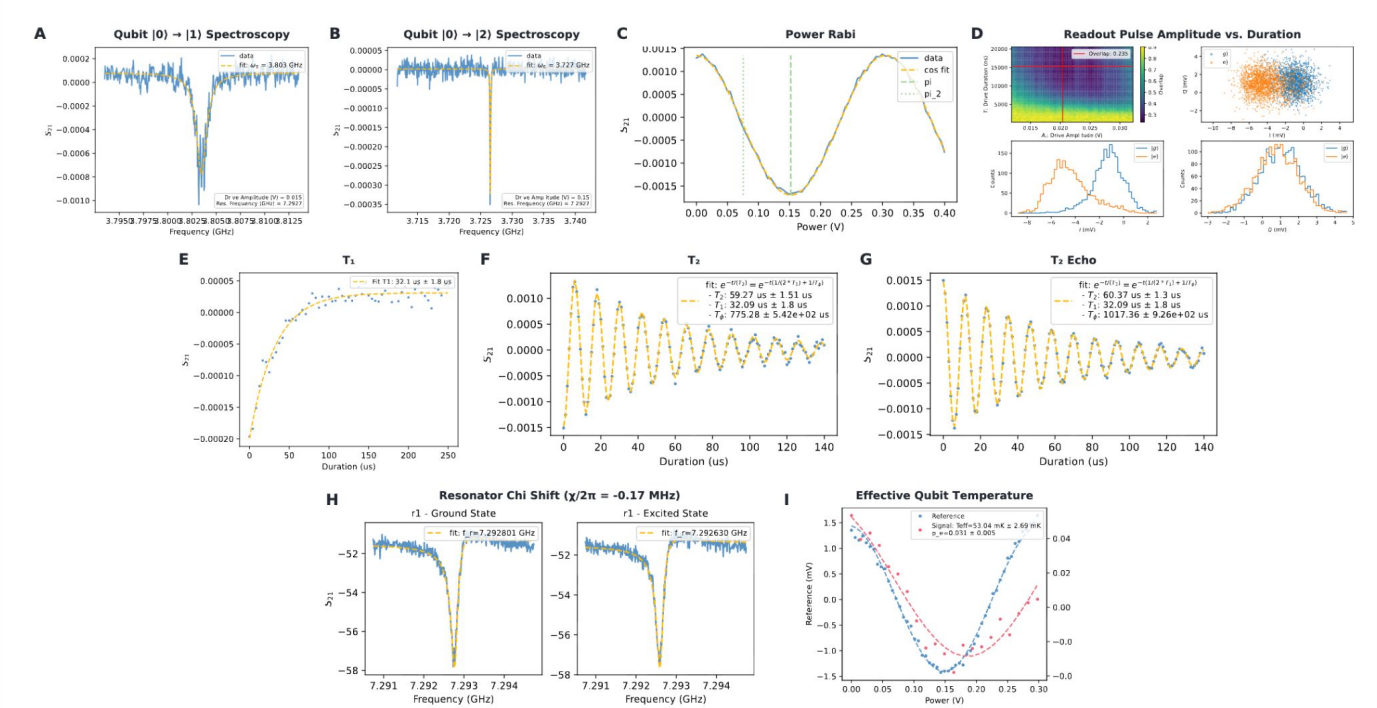
- MIT Engineering Quantum Systems (EQuS) Group led by Will Oliver
    - **GPT-5.6 Sol**, Ultra reasoning level, conducts superconducting-qubit measurements end to end
    - **Controls live laboratory hardware**, interprets data, and adapts subsequent measurement
- **Automates measurements** that can take a highly experienced researcher days



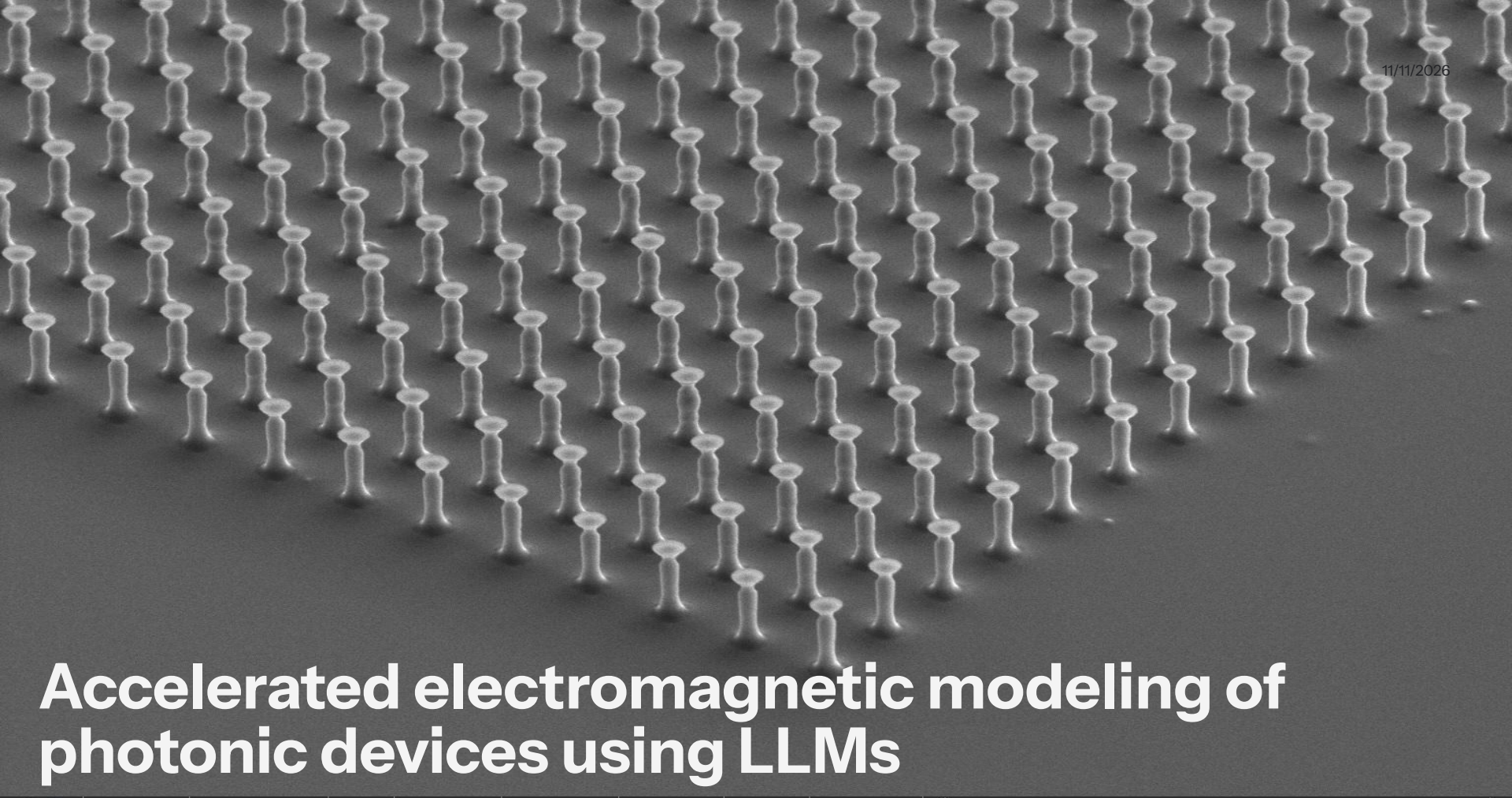


OpenAI. (2026, September 8). *How GPT-5.6 Sol helps run quantum computing experiments*. https://openai.com/index/codex-quantum-computing-experiments/

Autonomous Measurement Examples

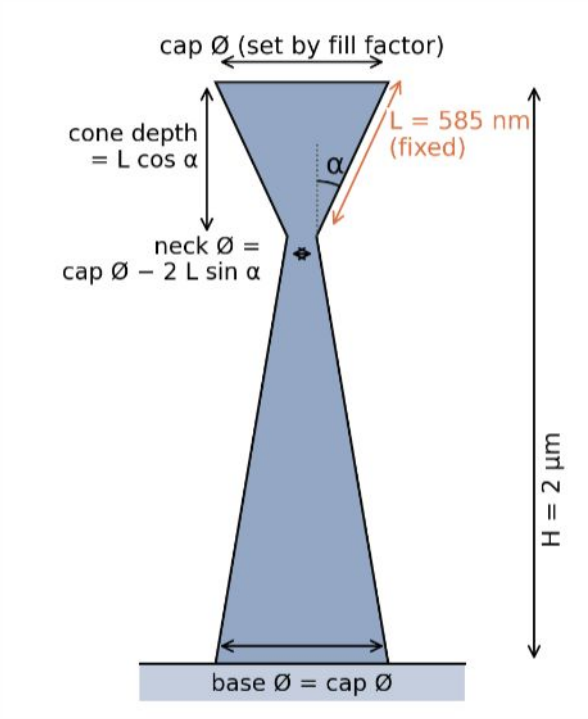


OpenAI. (2026, September 8). *How GPT-5.6 Sol helps run quantum computing experiments*. https://openai.com/index/codex-quantum-computing-experiments/



Accelerated electromagnetic modeling of photonic devices using LLMs

Accelerating Experimental Models in R-Lab

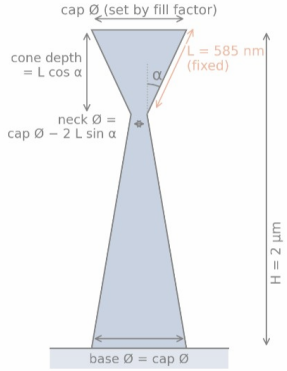


Initial proposed model

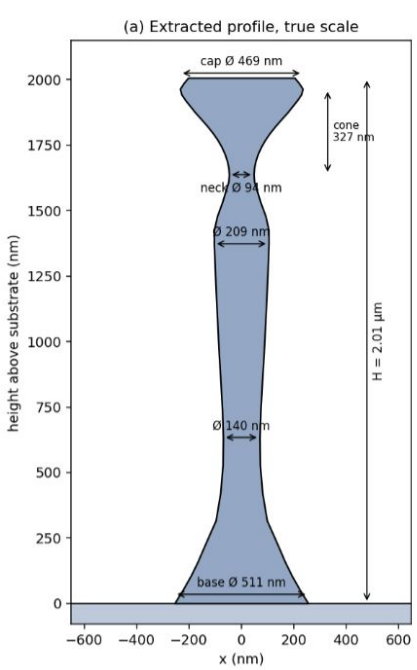
Accelerated electromagnetic modeling of photonic devices using LLMs

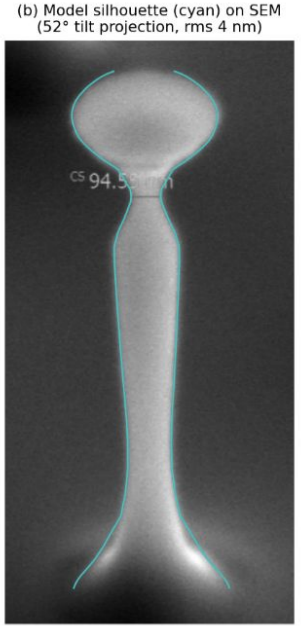
LLM generated based on experimental results

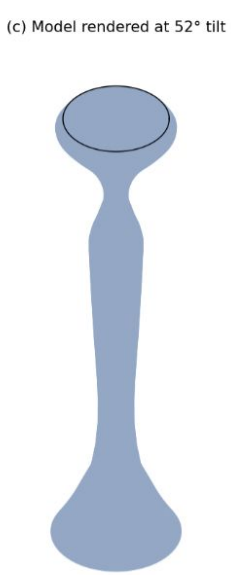
Accelerating Experimental Models in R-Lab



Initial proposed model







LLM generated based on experimental results

AI in Theory, Data Analysis, and Programming

Remi & Jordan

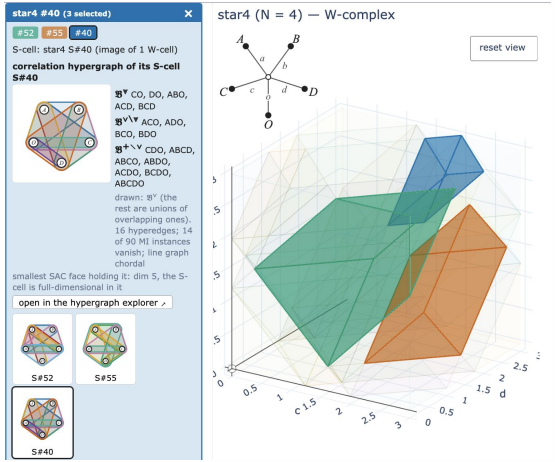
Theory

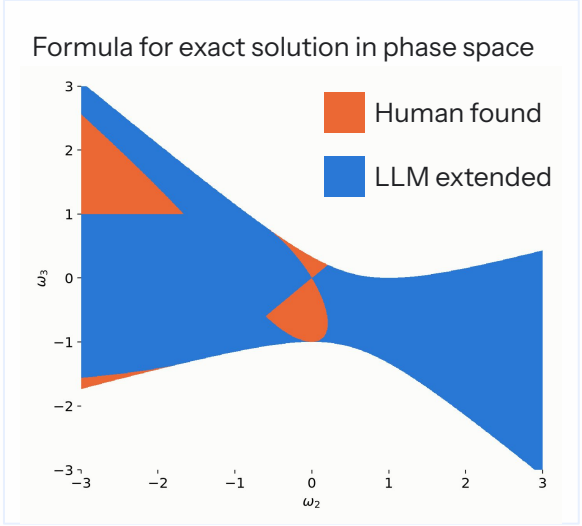
- LLMs can now lower the barrier for testing new ideas
  - Heavy calculations
  - Numerical checks, simulations
  - Programming, interactive visualizations
  - Conjectures extrapolating patterns

- Examples in High Energy Theory:

  - Graph minimum-cut structures
    - LLM-developed interactive visualization of higher-dimensional polytopes with Fable 5.1

  - Scattering amplitudes of water waves
    - Discovery of general formula using Opus 4.6

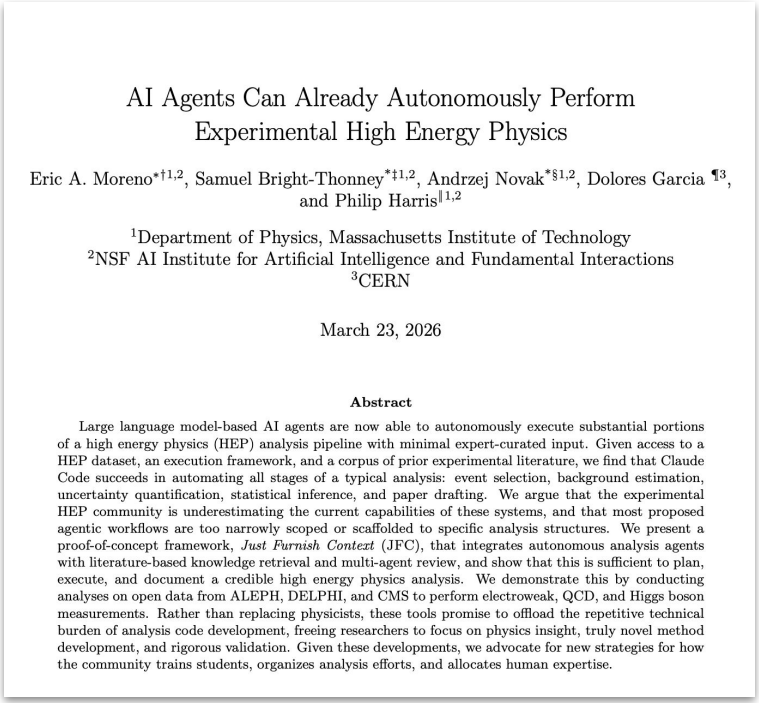




Surface Water Wave Scattering and the Hydrotope [arXiv:2606.28280]

Data Analysis

- Agentic coding can accelerate writing code, debugging, etc.
- Possible future (present?) for HEP-EX
  - JFC (Just Furnish Context) method
  - Researcher's job shifts from the mechanics to proposing questions and evaluating analysis quality



AI Agents Can Already Autonomously Perform Experimental High Energy Physics

Eric A. Moreno*†[1,2], Samuel Bright-Thonney*‡[1,2], Andrzej Novak*§[1,2], Dolores Garcia ¶[3], and Philip Harris‖[1,2]

[1]Department of Physics, Massachusetts Institute of Technology
[2]NSF AI Institute for Artificial Intelligence and Fundamental Interactions
[3]CERN

March 23, 2026

Abstract

Large language model-based AI agents are now able to autonomously execute substantial portions of a high energy physics (HEP) analysis pipeline with minimal expert-curated input. Given access to a HEP dataset, an execution framework, and a corpus of prior experimental literature, we find that Claude Code succeeds in automating all stages of a typical analysis: event selection, background estimation, uncertainty quantification, statistical inference, and paper drafting. We argue that the experimental HEP community is underestimating the current capabilities of these systems, and that most proposed agentic workflows are too narrowly scoped or scaffolded to specific analysis structures. We present a proof-of-concept framework, *Just Furnish Context* (JFC), that integrates autonomous analysis agents with literature-based knowledge retrieval and multi-agent review, and show that this is sufficient to plan, execute, and document a credible high energy physics analysis. We demonstrate this by conducting analyses on open data from ALEPH, DELPHI, and CMS to perform electroweak, QCD, and Higgs boson measurements. Rather than replacing physicists, these tools promise to offload the repetitive technical burden of analysis code development, freeing researchers to focus on physics insight, truly novel method development, and rigorous validation. Given these developments, we advocate for new strategies for how the community trains students, organizes analysis efforts, and allocates human expertise.

AI Agents Can Already Autonomously Perform Experimental High Energy Physics [arXiv:2603.20179]

Research

Jordan

Research expectations

The responsibility of a graduate student is to learn how to produce *future* research given the tools that *will* exist.

Avoid using ‘results’ as a proxy for grad student progress

Want to avoid:
“Let’s use LLMs to publish papers faster than we can understand/learn from them.”

Responsibility for LLM generated text

Want to avoid:
“Can you read this 100-page AI-generated document to see if there’s anything useful?”

Using LLMs to enhance skills without losing them

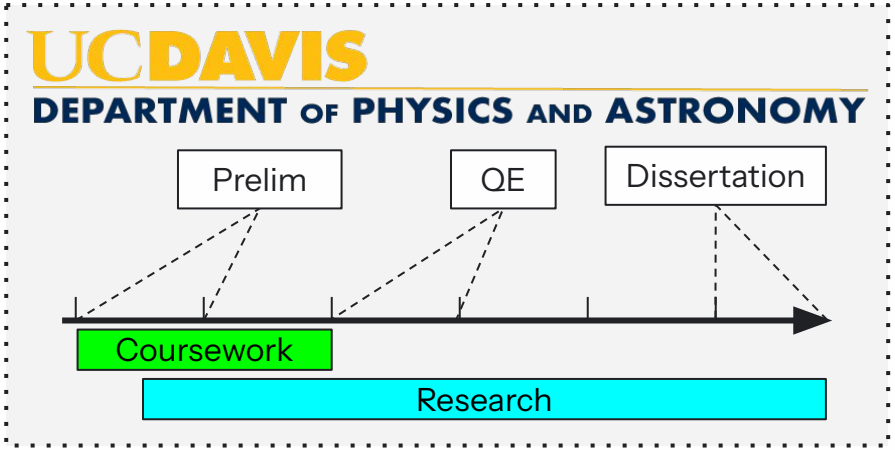
Want to avoid:
“Why even bother with LLMs?”
“Why even bother with learning anything?”

On the graduate program

- What are the enduring skill sets?
  - Abstract problem solving
  - Physical intuition
  - Research taste

- Guidelines and regulations
  - LLM use for QE or Dissertation?
  - AI disclosure policy for written work

- Advisors and grads should stay informed about LLM impact and capabilities

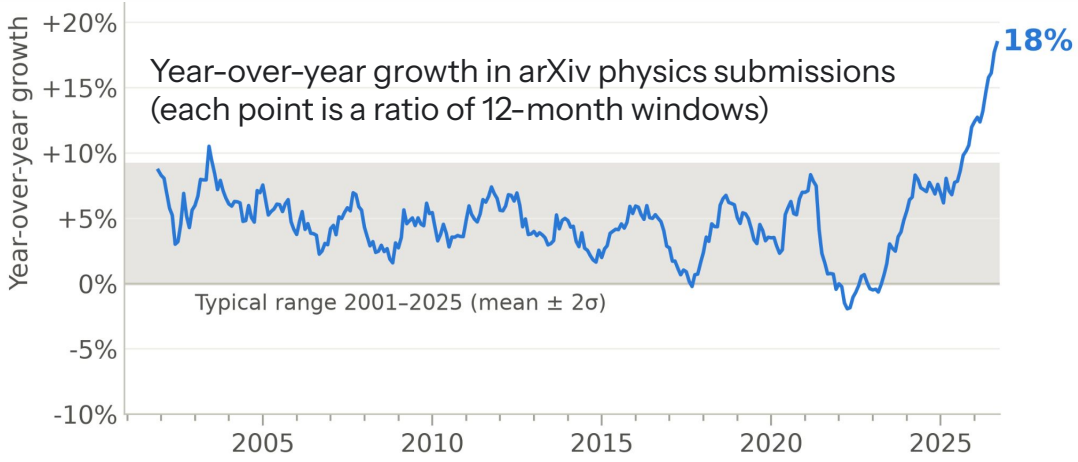


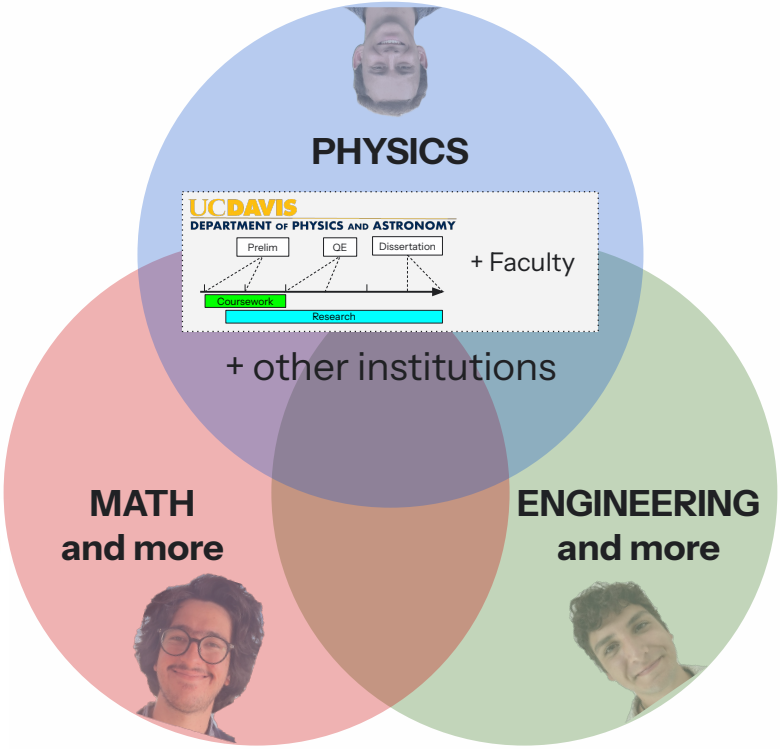
We should specify the department's guidelines in writing.

https://physics.sf.ucdavis.edu/sites/g/files/dgvnsk16581/files/media/documents/Graduate_Student_Timeline.pdf

On research at large

- LLMs have present shortcomings
  - Inadequate citations, incomplete justifications, over-generalizations, …
- Publication numbers growing fast
  - Peer review slower, sifting through literature, reproducibility, …
- Learn from/work with adjacent fields





Data gathered from *https://arxiv.org/year/<archive>/<year>* summing the 12 physics archives
See also [2610.34449v1]

Learning and Teaching

Remi

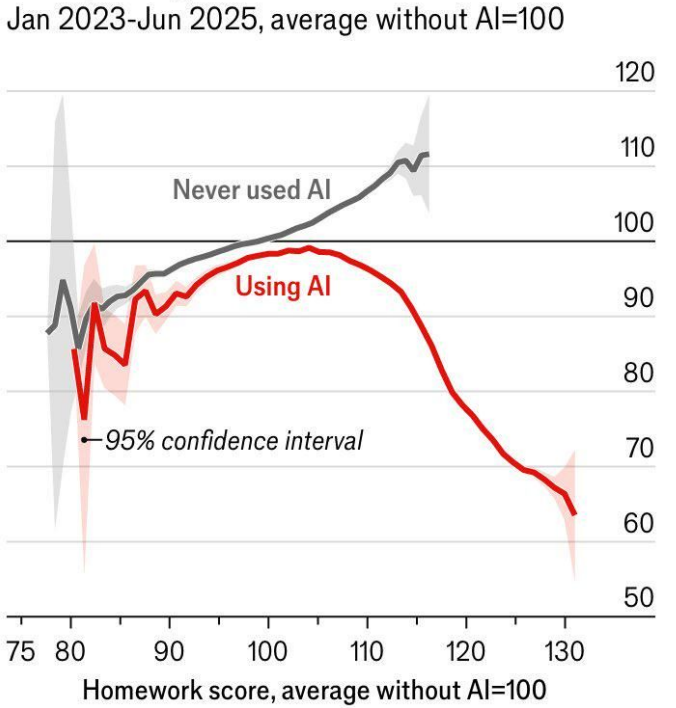


How education is evolving

- Instructors already dealing with effects of student LLM use
  - Pros: patient tutor, easier to make course materials
  - Cons: LLMs can solve traditional problem sets easily
  - **“Cognitive offloading”**

- TA observations: basic skills in lower-division classes lagging behind, AI may be one cause

If LLMs can solve physics problems better than humans, what is the point of teaching physics classes?



*26,811 pupils aged 12-18
Source: “The generative AI learning penalty: evidence from Chinese secondary education”, by D. Strömberg et al., preprint June 2026

Stromberg, David and Lei, Victor and Wu, Yanhui, The Generative AI Learning Penalty: Evidence from Chinese Secondary Education (June 02, 2026). SSRN, graphic from The Economist.

Purpose of teaching

The primary goal of physics coursework at any level should be the development of abstract problem-solving skills and physical intuition.

Emphasizing problem-solving skills over memorization of solutions

Want to avoid: Students memorizing a particular solution, without developing understanding that generalizes.

Encouraging responsible LLM usage by both students and instructors

Want to avoid: Students copying LLM written problem solutions; instructors AI-generating all of their course content.

Prioritizing teacher-student and student-student interactions

Want to avoid: Students using LLMs as their sole source of truth and not interacting with teachers and peers

Changes in teaching

Students will use LLMs for better or worse. We must be bold and adapt.

- Instructors should restructure courses and update teaching methods
  - Shift towards in-person assessment (without placing the whole burden on TAs)
  - Require clear AI usage policies for every course in syllabi
  - No required LLM usage in courses

- Department should draft policy on AI best practices and norms (e.g., instructor usage, student privacy, etc.)

Faculty and graduate students should hold meetings—we interact the most with students, we want to be involved!

[MIT Report] https://bpb-us-e1.wpmucdn.com/sites.mit.edu/dist/d/2418/files/2026/09/AI-Committee-Final-Report-Aug-13.pdf

Procurement

Graduate students should have equitable access to LLMs via an agentic platform to support scientific use.

- Concrete suggestion : use one of the ANTHROPIC plans to provide ‘standard seats’
  - Some groups are already on a “Claude Team plan for scientists”
    (temporary promotional pricing $0/month for a standard seat)
  - We could seek a “Claude for higher education” enterprise plan
    ($20/month per standard seat)

- Future discussions should consider local alternatives
  - Data security and sovereignty (FERPA compliance, sensitive research, …)
  - Thoughtful experimentation for physics-specific use cases

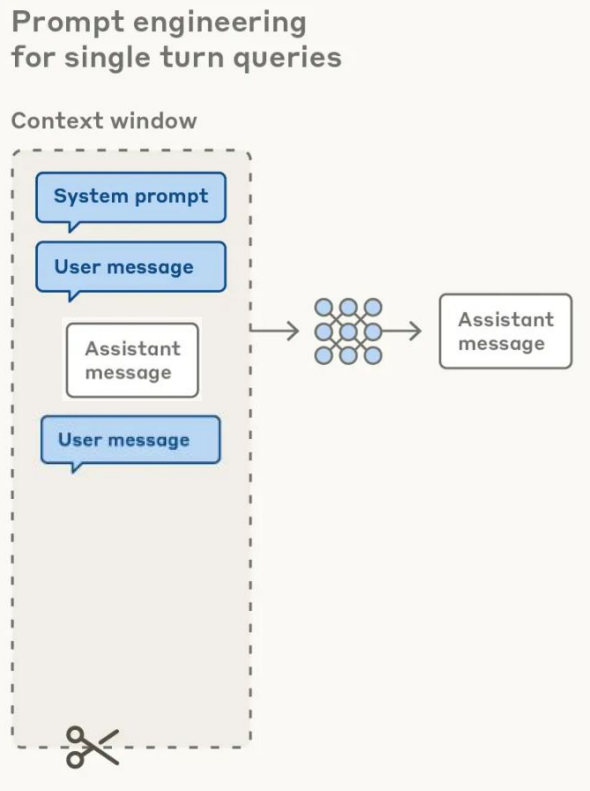
https://support.claude.com/en/articles/16634237-claude-team-plan-for-scientists
https://claude.com/contact-sales

In Conclusion...

We must hold regular meetings between grads and faculty to adapt to the changes in science and education.

Thank you!

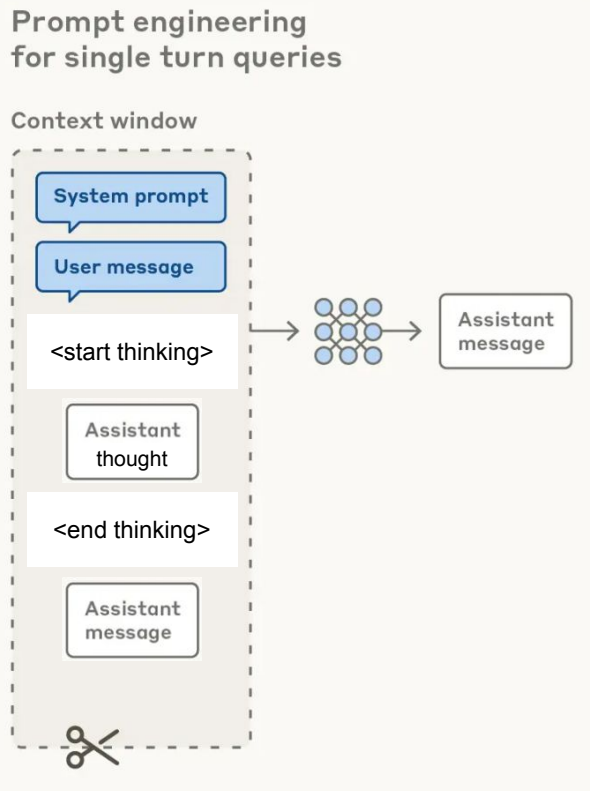
2022 - GPT3.5 is released by OpenAI



GPT3.5 is trained via fine-tuning on human text.
It is a ‘next-token predictor’.

https://www.anthropic.com/research/effective-context-engineering-for-ai-agents

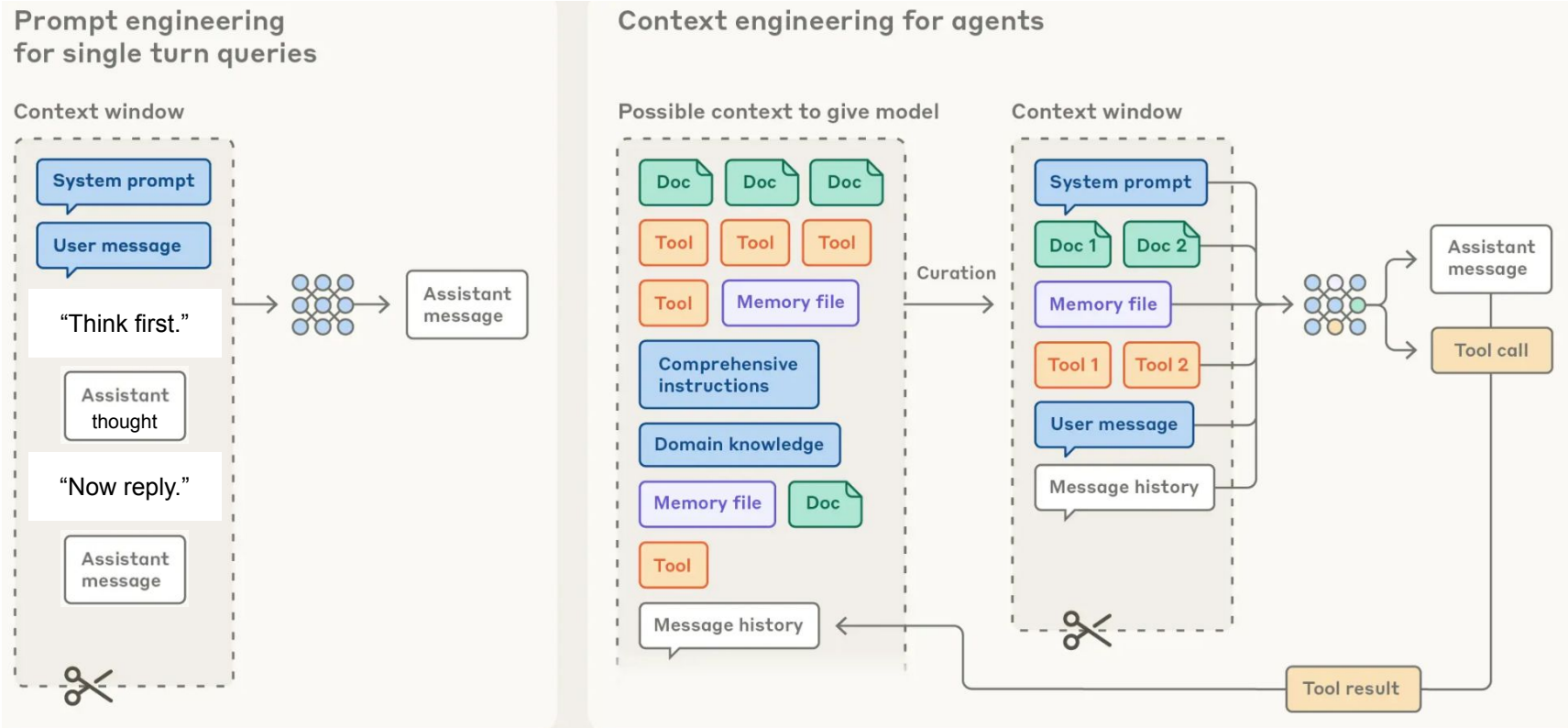
2024 - OpenAI o1 is released by OpenAI



After fine-tuning,
the model is reinforced to do well on tasks.
It 'thinks out loud to itself' before speaking.

https://www.anthropic.com/research/effective-context-engineering-for-ai-agents

2025 - ‘Agentic’ models curate context and act independently



https://www.anthropic.com/research/effective-context-engineering-for-ai-agents