Physics Grad LLM Statement

"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[1], Yale Physics Chair

> *LLMs and Physics:* Large Language Models and the infrastructure around them are developing rapidly[2], with them having grown from search assistants to being capable of meaningful problem solving in theoretical and experimental domains[3]. We must consider carefully how the Physics program evolves in response to this circumstance.

➔ *Looking around:* UC Davis[4], other universities[5], arXiv[6], and journals[7] are reacting.

Preparing discussions regarding LLM use

Given how rapidly the technology develops around us, and the diversity of opinions surrounding it, we are unlikely to land upon a complete solution in one statement, retreat, or entire quarter. The contents of this document are only recommendations, and there are many questions students and faculty may have for each other, which we should plan to discuss again.

> *General Forum:* As soon as possible, we should hold short, regular, structured open meetings[8] to discuss and make decisions on AI-related topics affecting the graduate program's operations, and the impacts on the physics community as a whole.

➔ *Teaching-specific:* A substantial fraction of changes are teaching-specific. Instructors and graduate students should collaborate on changing how students are evaluated, informing students and faculty about AI basics and norms, and comparing outcomes[9].

Research and the relationship to one's research advisor

One of the foremost goals of a graduate program is to prepare graduate students to perform meaningful research. Research itself may become a collaboration between human insight and machine-assisted work.

> *Expectations:* A graduate student is expected to be trained to produce excellent future research using the tools that will exist; producing present research is valuable, but it should not be the sole expectation. A research advisor is expected to assist in this training by helping the graduate student learn the scientific context and norms. This job requires research advisors to engage with how LLMs are affecting their field.

➔ *Research skills:* As LLMs become increasingly capable of the skills in a graduate student's toolkit, there are increasing incentives to use them to replace parts of one's work. We should prepare guidelines stating which skills should nonetheless be practiced by graduate students.

---

[1] Quote taken from In an Age of AI, a Physicist Seeks What Endures | Quanta Magazine.
[2] Though somewhat dated, one might consider this pair of posts by METR: Time Horizons and Across Domains.
[3] See the case studies of [2610.17193] and the [2610.16733] TheoPhysEval benchmark. The former also includes guiding principles.
[4] See the Advancing AI at UC Davis report by the UC Davis AI Steering Committee.
[5] See for example the Report of MIT's Ad Hoc Committee on AI Use in Teaching, Learning, and Research Training.
[6] See arXiv's blog post: Fair Moderation, Equitable Access, and AI: arXiv's Updated Rate Limit Policy.
[7] See for example Nature's editorial policy: Artificial Intelligence (AI) | Nature Portfolio.
[8] A possible option would be to have 2-hour twice-quarterly meetings, such that the meeting is open to anyone to attend, but with short presentations / an agenda selected by a committee of faculty and graduate students beforehand.
[9] In particular, we should implement standardized metrics that track which LLM-enhancements affect courses well.

Guidelines and regulations: We should determine guidelines and regulations[10] indicating how both intermediate steps of research, and research-related writing, are to be handled. In particular, there should be standards for AI-use disclosures, and researchers must be accountable for the content of their writing, regardless of how it is produced.

➔ *Research assessment:* A graduate student's research-related progress is assessed in the Qualification Examination and in the Thesis. There should be clear regulations regarding AI use for both of these far in advance of any student's assessment.

Learning and Teaching

Aside from doing research, a graduate student engages with their own coursework and teaching duties. For both graduate and undergraduate education to be successful, existing teaching and learning methods should be reconsidered and updated where necessary.

Instructor/TA responsibility: Inevitably, students will use LLMs for their coursework, for better or worse. Instructors should restructure their courses[11] to take this into account, without losing sight of the purpose of learning physics.

➔ *Purpose*: The primary goal of physics coursework at any level should be the development of abstract problem-solving skills and physical intuition. These are more likely to be transferable to a world with LLMs, unlike the acquisition of memorized knowledge or routine computational skills.

➔ *AI Policy*: In each course, instructors should clearly communicate how AI can be used responsibly by students[12] and how assessments might depend on it.

➔ *Grading*: Courses should rely more on in-person evaluations as well as interactions with TAs[13]. At this time, we believe that LLM usage should not be required in any course.

Procurement of LLM access

We believe that interested graduate students should have the ability to access enterprise models with reasonable scientific capabilities. In the interest of making a concrete suggestion:

Anthropic: Any graduate student interested in having access to an enterprise LLM should be able to acquire access via a 'standard seat' on a plan with Anthropic[14].

➔ *Other providers:* Individual groups may consider paying for other models. We discourage their endorsement due lesser commitments to scientific integrity[15] and AI safety[16].

➔ *Local Alternatives:* We should consider access to open-weight models[17] to maintain independence from technology companies and thoughtfully experiment with LLMs.

---

[10] For example, the APS has an updated policy, allowing "more substantive AI use … including tasks like literature synthesis, data analysis, scientific reasoning, figure generation, translation, and more — with disclosure."

[11] We should decide carefully how LLMs are used by instructors in this restructuring. In some cases, LLMs can be helpful, such as to produce teaching materials. In other cases, LLMs must be avoided, such as in cases where student data must be kept private.

[12] One easy-to-implement idea is to provide students with a department-standard or course-standard Markdown file containing an extended LLM prompt, so that the LLM acts like a responsible tutor, guiding the student without revealing solutions.

[13] For example, there have recently been experiments in graduate courses with weekly one-on-one TA-student meetings, whose purpose is to ascertain each student's understanding of submitted homework assignments.

[14] The Claude Team plan for scientists is $0/month for a standard seat. Claude for Education does not have published figures, but Anthropic's "buying agent" quotes $20/month. We recommend the Team plan for its lower price and greater flexibility.

[15] This is hard to judge as science is both a genuine return from technology, and a marketing tool for AI companies, c.f. [Science] | OpenAI , [Science] | Anthropic. OpenAI's release of mathematical results has been described by some mathematicians as "total disregard for the norms of scientific research".

[16] The Future of Life Institute provides a very coarse-grained look at transparency and safety.

[17] See for example the Open Math Model for what such experimenting can look like at scale.