Last updated on April 29, 2026

INTERIM Policy on Generative/Agentic Artificial Intelligence (AI) Usage for Graduate Students

Department of Psychology, University of Maryland

Purpose

Artificial Intelligence (AI) has changed the graduate education landscape from course instruction to research practices. The purpose of this document is to provide information on the Department of Psychology's Generative Artificial Intelligence (AI) Usage Policy for Graduate Students. Our goal is to provide responsible guidelines that align with the department's learning and developmental goals for graduate students, such as gaining a breadth of knowledge, developing independent research skills, and growing one's expertise. We also acknowledge the growing ubiquity and utility of AI in academic contexts and beyond. As the abilities of generative/agentic AI change, so may this document and associated policies. This current version reflects our best assessment of the current landscape of generative/agentic AI.

Policies apply to any student registered as a PhD or MPS student in the Department of Psychology, as well as any UMD graduate student who is taking a graduate-level Psychology course, collaborating with Psychology faculty on research, and/or working as a teaching assistant in a Psychology course. *For clinical students, you should also review the AI policy for the UMD Psychology Clinic (INSERT HERE).

The document should be used in conjunction with the University of Maryland's Guidelines for AI Use. This set of guidelines encourages clarifying AI permissions and restrictions with one's instructor, mentor, etc.. Students should also be aware of APA's policy on generative AI.

General Considerations for AI Use

Though AI can be used to increase efficiency and personalize the learning experience, it can also produce information that is inaccurate and presents a risk of violating privacy, copyright, and ethical guidelines. Before discussing the policy, we present an overview

of some of the nuanced considerations in using AI during your graduate education years and beyond[1].

There are many opportunities for AI application and integration in graduate education. Generative/agentic AI programs are tools that can drastically accelerate coding, assist in literature searches, and refine ideas. Further, they could be used to help scaffold certain types of skills (e.g., serving as a writing coach or a stats tutor) for students who enter graduate school with less experience or other barriers. We also acknowledge that AI is here and ubiquitous. Some students will plausibly co-author, critique, and propose experiments with AI tools in the near future. Graduates who have no experience using AI may enter a workforce expecting fluent, critical human–AI collaboration.

However, there are risks of irresponsible AI use and AI overuse, particularly for junior scholars in training. Using AI without attribution and without accuracy checks could result in running afoul of copyright laws and plagiarism rules (if not the letter, then the spirit), as well as producing and disseminating work that includes false citations and erroneous interpretations. Further and most importantly, overreliance on AI could prevent skills from developing or atrophy existing skills. Preliminary evidence suggests that frequent AI use, cognitive off-loading (i.e., using AI to replace deep thinking), and confidence in AI's abilities is associated with lower levels of self-reported critical thinking (Gerlich, 2025; Lee et al., 2025). Our graduate program values deep, independent thinking from students and overuse of AI would undermine this key developmental goal.

Overall, use of AI in graduate work (when permitted) should be thoughtful and in service of the student's development as a scholar. The risks necessitate some restrictions, outlined in the rest of the document.

Definitions

Generative AI. UMD defines Generative AI as "a subset of artificial intelligence that focuses on producing novel content by learning patterns from training data" (Generative AI at UMD, n.d.). *Agentic AI* is an advanced subtype of AI that can "complete tasks independently or with minimal human supervision (Stackpole, 2026). Commonly used AI tools include ChatGPT, Google Gemini, Anthropic's Claude, and Microsoft Copilot.

---

[1] In aligning with the policy on disclosure (see point 2 in the policy section), the "General Considerations for AI Use" section utilized Claude Opus 4.7K and guided prompts provided by A.J. Shackman. The section was written and checked for accuracy by J. Wessel, integrating some of the AI output with faculty feedback and alternative sources. Full prompts and associated outputs are in Appendix A at the end of this document. It's a great read with a lot of extra information – I recommend it! (J.W.)

Last updated on April 29, 2026

UMD also has its own chatbot, TerpAI. We are not considering integrated grammar checks (e.g., Grammarly) as a generative AI tool for this policy's purposes.

Policy. The policy presented in this document refers to rules for maintaining research, instructional, and/or other academic integrity. Violating said rules could result in a referral to the Office of Student Conduct for the student (process described here).

Guidelines. Guidelines in this document are presented in the form of checklists to discuss with one's advisor/supervisor/instructor, depending on the context. The checklists serve as a guide for discussion and mutual understanding of the specific AI policy for that context, but they are not policies on their own. Each instructor/advisor/supervisor can create their own classroom/lab policy based on their specific needs, norms, and requirements. The guidelines are intended to be a resource for both students and faculty to avoid misunderstanding and promote academic and research integrity. In this vein, classroom and lab policies and individual use should be guided and informed by other relevant university policies and guidelines, including those provided by the Office of Research Integrity.

AI Department Policy

Generative AI may be appropriate to use as a tool in certain academic and research contexts, but there are certain **general rules** that we require students to abide by in order to align with university and field standards for academic integrity, as well as to maintain our high standards for student learning and development:

1. Students and advisors/instructors/supervisors should discuss rules around generative and agentic AI use in advance of (or at the very beginning of) a given project/course. Given constant advances with AI, the student is responsible for having ongoing discussions with their advisor/instructor/supervisor on how they are using AI on their respective project.
2. Any use of generative/agentic AI produced text (including code), materials, and/or images must be disclosed in the text, presentation, or other medium within which it was used. Use of generative AI to analyze data should also be disclosed. Transparency of AI includes which program was used (model name and version/number), how it was used (e.g., creating text, creating code), the prompts that were used, and which exact portions of output are attributable to generative AI. Students should disclose in an appendix, a footnote, or

supplementary materials as a statement, as seen in this example, adapted from Kari Weaver, University of Waterloo:

AID Statement Example

"Artificial Intelligence Tool: ChatGPT v.4o and Microsoft Copilot (University of Waterloo institutional instance); *Data Collection Methods:* ChatGPT was used to create the first draft of the survey instrument; *Data Analysis*: Microsoft Copilot was used to verify identified themes coded from open ended survey responses; *Privacy and Security*: no identifiable data was shared with ChatGPT during the design of this study, only the University of Waterloo institutional instance of Microsoft Copilot was used to analyze any anonymized research data in compliance with University of Waterloo privacy and security policies; *Writing—Review & Editing*: ChatGPT was used in the literature review to provide sentence-level revisions"

3. Any use of generative/agentic AI must maintain others' privacy. It is prohibited to upload others' materials (e.g., student papers, client transcripts, instructor slides, others' research articles) into generative AI tools without the original authors' permission. Exceptions are materials that are already in the public domain, such as open access papers that do not require permission to use. It is also prohibited to upload personally-identifying information of study participants, students, or clients.
4. Any use of generative/agentic AI must be checked by the student for accuracy. This includes ensuring that sources are accurate in terms of details and attributed claims, that code functions accurately (not just "runs", but each line is constructed as intended), etc.. AI summaries can lead to individuals not thoroughly reading cited sources, potentially leading to misinterpretation and errors. Here are two resources outlining best practices for citing sources within sources (ORI guidelines; Illinois Tech Guidelines). We suggest including a brief overview of the accuracy check process at the end of the AI disclosure statement. An example:

Accuracy Check: All output (sources, claims) was checked against primary sources by the first author for accuracy.

5. One major purpose of theses and dissertation work is the development and demonstration of independent thinking skills and expertise. Further, the Graduate School's rule restricting use of copyrighted and/or plagiarized material

in theses and dissertations may be violated by overreliance of AI-generated text. As such, we require first drafts of theses and dissertations to be generated by the student without generative/agentic AI assistance. Put in another way, original writing for your thesis/dissertation should be your own and should not "plagiarize" or copy/paste from AI programs. Other restrictions around AI use (e.g., for help with idea generation, writing revisions, editing suggestions, coding, etc) should be discussed with one's thesis/dissertation chair. If a student has co-chairs, that advising team will need to decide jointly what is allowed. As with any aspect of a thesis or dissertation, committee members may have different beliefs about appropriate AI use. Permitted AI use should be discussed and agreed upon (ideally prior to or) during the proposal meeting.

6. Comprehensive exams differ in format and process by program area, but generally aim to assess student mastery and in-depth understanding of their discipline. As such, we believe there should be restrictions on AI use during comprehensive exams. The exact restrictions will be determined by program area faculty and communicated to students via area handbook and exam instruction documents.
7. Creating and clearly communicating context-specific AI rules is the responsibility of the advisor/instructor/supervisor. Adherence to general generative AI rules and context-specific rules is the responsibility of the graduate student. Proclaimed lack of awareness is not an excuse for violating rules. If you are uncertain, ask first!

Guidelines for Specific Contexts

Our department includes many courses and research labs, which may have their own specific norms and rules around AI use. The checklists below are guides for beginning the conversation around generative AI use in a classroom, lab, or other context. The guidelines do not include the general rules (see previous section), **which apply in every departmental context.** Not all items will be relevant to every context. Context-specific rules that go beyond the general rules should be viewed as additional policies to follow for that particular course/research project/etc.

TA assignment guidelines. To be decided by the supervising instructor.

Can I use generative AI to assist in…	Yes/No/Specific Restrictions

1. Creating a grading rubric?	
2. Editing student work?	
3. Coming up with ideas or creating materials for class activities?	
4. Coming up with lecture materials (e.g., slides)?	
5. Coming up with lab materials?	
6. Creating emails to respond to student or instructor questions?	
7. Creating exam and quiz items?	
8. Creating homework items or problem sets?	
Other permissions or restrictions:	

Research assistant or collaboration guidelines. To be decided by the faculty advisor or leading PI on the research project and/or grant.

Can I use generative AI to assist in…	Yes/No/Specific Restrictions
1. Brainstorming research topics?	
2. Conducting a literature review?	

3. Assistance with interpreting analytic results or choosing appropriate statistical tools?	
4. Creating code for data analysis?	
5. Checking code?	
6. Creating materials for experiments, surveys, interventions, etc.?	
7. Writing an article or conference submission?	
8. Developing an outline for an article or conference submission?	
9. Writing a grant proposal?	
10. Improving writing quality (e.g., asking for suggestions to improve already-written text from the student)?	
11. Formatting tables or figures?	
12. Formatting reference sections?	
13. Creating slides for conference/lab/other presentations?	
14. Creating project meeting agendas?	
Other permissions or restrictions:	

Last updated on April 29, 2026

Coursework. To be decided by the instructor of the graduate course.

Can I use generative AI to assist in…	Yes/No/Specific Restrictions
1. Brainstorming ideas for a class paper or other submitted work?	
2. Conducting a literature review for submitted work?	
3. Creating a summary of an assigned article?	
4. Developing an outline for submitted work?	
5. Creating a summary or practice exam/quiz from instructor recorded lectures or uploaded slides? *Note - this violates copyright laws if done without permission*	
6. Creating code used in submitted work?	
7. Writing answers to exams?	
8. Writing answers to non-exam submitted work?	
9. Improving writing quality (e.g., asking for suggestions to improve already-written text from the student)?	
10. Formatting tables or figures for submitted work?	

11. Formatting reference sections for submitted work?	
12. Brainstorming ideas and/or questions for leading class discussions?	
13. Creating slides for in-class presentations?	
Other permissions or restrictions:	

Comprehensive exam guidelines. To be decided by area program.

Can I use generative AI to assist in…	Yes/No/Specific Restrictions
1. Brainstorming ideas?	
2. Conducting a literature review?	
3. Assistance with interpreting analytic results or choosing appropriate statistical tools?	
4. Creating code for data analysis?	
5. Checking code?	
6. Improving writing quality (e.g., asking for suggestions to improve already-written text from the student)?	
7. Formatting tables or figures?	
8. Formatting reference sections?	
9. Creating table, figures, or other supplementary materials?	

Other permissions or restrictions:	

Thesis/dissertation guidelines. To be decided by committee chair(s).

Can I use generative AI to assist in…	Yes/No/Specific Restrictions
1. Brainstorming ideas?	
2. Conducting a literature review?	
3. Assistance with interpreting analytic results or choosing appropriate statistical tools?	
4. Creating code for data analysis?	
5. Checking code?	
6. Creating materials for experiments, surveys, interventions, etc.?	
7. Improving writing quality (e.g., asking for suggestions to improve already-written text from the student)?	
8. Formatting tables or figures?	
9. Formatting reference sections?	
10. Creating proposal/defense slides?	

11. Compiling committee feedback?	
12. Crafting responses to committee feedback?	
Other permissions or restrictions:	

Last updated on April 29, 2026

Appendix A

Claude, could you please write a "high-level", executive summary, of the potential pros and cons of embedding generative AI/LLM's (e.g. Claude Opus 4.7) in doctoral training in neuroscience and psychological science (i.e. fundamental "blue-sky" science training at he PhD level)? In writing the paragraph, you should limit yourself to no more than 1,000 words. You should also consider how the workforce and AI capabilities will change over the careers of current students. How should we optimally prepare them? How can cultivate strong intellectual habits and side-step the erosion of foundational intellectual skills?

=== Process: How We Got to the Specifics of the Prompt

Before I draft this, a few questions and observations that might sharpen the result:

Clarifying questions

1. **Audience and venue.** Is this for a departmental committee, a funding body (NIH, NSF, Wellcome, ERC), a graduate school steering group, or internal circulation among faculty? The register and which concerns get foregrounded (training grants vs. pedagogy vs. research integrity) shift noticeably across these.
2. **Stance.** Do you want a genuinely balanced pros/cons synthesis, or do you have a leaning — e.g., you're broadly pro-integration but want risks surfaced honestly, or you're skeptical and want the strongest case for adoption fairly represented? I can write either, but "executive summary" often implies a recommendation, not just a ledger.
3. **Scope of "embedding."** Embedding could mean anything from (a) permitting use with disclosure, to (b) formal curriculum modules on AI literacy, to (c) AI as a co-pilot for coding/writing/literature review, to (d) AI as a reasoning partner in hypothesis generation and experimental design. The cons look very different across these levels.
4. **Neuroscience + psych science together, or separate?** They share a lot but diverge on things like wet-lab vs. computational training, clinical/human-subjects considerations, and the replication-crisis backdrop in psychology specifically. Want me to treat them jointly or flag where they differ?

Suggestions for refinement

Last updated on April 29, 2026

- **1,000 words is generous for a true executive summary.** Most execs want 250–400. If this is genuinely exec-facing, I'd suggest ~400 words with an optional longer appendix. If it's more of a white-paper-style briefing, 1,000 is fine. Which is it?
- **The "career horizon" framing is the most interesting part of your prompt** and the part most briefings get wrong. A PhD starting now finishes around 2031 and hits tenure review around 2038–2040. The honest answer is that nobody knows what AI capabilities or the academic/industry job market look like then — any summary that pretends otherwise is performing certainty it doesn't have. I'd recommend the piece explicitly name this uncertainty rather than paper over it with confident predictions. Want me to take that approach?
- **"Erosion of foundational skills" deserves more precision than it usually gets.** It's worth distinguishing skills that are genuinely at risk (deriving things from first principles, close reading, sitting with confusion, statistical intuition built through hand-calculation) from skills that were always partly outsourced anyway (looking up syntax, formatting citations). A good summary shouldn't treat all "foundational skills" as equally threatened or equally worth protecting.
- **One substantive point you might want addressed explicitly:** there's an emerging concern, with some empirical backing, that heavy LLM use during learning produces fluent-sounding outputs while hollowing out the underlying understanding — the student can produce the paper but can't defend it in depth. This matters enormously for PhD training specifically, where the *viva* / defense and the ability to think on your feet at conferences are the actual product. Want this foregrounded?
- **A missing dimension you might want in scope:** mentorship and the advisor-student relationship. LLMs can absorb a lot of the "dumb questions" students previously took to postdocs and PIs, which has upsides (less friction, faster iteration) and real downsides (erosion of the apprenticeship model that's historically how taste and judgment get transmitted in science).

If you'd rather I just proceed, my default would be: ~600 words, balanced but with a clear recommendation, treating neuro and psych jointly with flagged differences, explicitly naming career-horizon uncertainty, and covering the mentorship/apprenticeship angle. Say the word and I'll draft it — or answer any subset of the above and I'll tailor it.

Last updated on April 29, 2026

1. *Audience. A mix of Department and Graduate School faculty. The primary audience should be faculty in neuroscience and psychological science, but the verbiage may be circulated amongst faculty in other programs or at the overarching graduate school or campus level. 2. I appreciate your comment, but I would like a genuinely balanced "ledger," at least for this first round. 3. It is likely that the scope of embedding will vary across different laboratories and courses. It would be useful to write the summary paragraph without regard to specific embeddings. Please also include an Appendix that highlights pros and cons for specific use-cases/scenarios, of the kind you mentioned. 4. Treat them jointly, but add a second Appendix that flags potential differences across disciplines. In crafting the second Appendix, you should recognize that both disciplines cover many, varied sub-disciplines. For example, some neuroscientists primarily work with animal models, whereas others primarily work with fMRI or other neural recordings. In the latter case, much of the work is computational and analytic. 5. Please craft 2 versions of the Executive Summary. One should be 300 words or less. The second can be as long as 1,000 words. 6. Regarding "career horizon." Yes, you should acknowledge this uncertainty, but you can safely assume that future capabilities will at least be as good as the most sophisticated platforms that are currently commercially available (e.g. Opus 4.7 Extended). 7. Regarding "erosion." You make excellent points. For the 2 versions of the Executive Summary, I think you will have to be somewhat vague. However, I would like you to craft a third Appendix that delves more deeply into some specific skills and the degree to which they are threatened and/or worth protecting. Explore the tradeoffs. 8. Regarding "missing dimension." Please incorporate this into the 3rd appendix, where you explore some specific skills and tradeoffs. Does all of this make sense? Any concerns or suggestions?*

Yes, this all makes sense and is a clear brief. A few remaining questions and one suggestion before I draft:

Questions

1. **Length targets for the appendices.** You gave word counts for the two summary versions but not the appendices. My instinct is ~500–800 words per appendix, but if you have a preference — either tighter (so the whole document stays under, say, 3,500 words) or more expansive — tell me now. Appendix 3 in particular could easily run long given the skills inventory.
2. **Empirical grounding.** Should I cite the emerging literature on LLMs and learning (e.g., work on metacognitive offloading, the MIT "Your Brain on

ChatGPT" study, recent pieces in *Nature* on AI in science training), or keep it as a reasoned-from-first-principles briefing? Citations add credibility with this audience but also add length and a maintenance burden if the document circulates for a while. I'd lean toward a handful of well-chosen citations in Appendix 3 only, but happy to do none or more.

3. **Appendix 3 — descriptive or prescriptive?** When I dig into specific threatened skills, do you want me to stop at "here's the tradeoff" or go further into "here's how you might structure training to protect X" (e.g., AI-free qualifying exams, first-year methods bootcamps where hand-calculation is required, oral defense components designed to catch hollow understanding)? The prescriptive version is more useful but also more opinionated — which may or may not be what you want for a balanced ledger.
4. **Tool-naming.** Should I name specific platforms (Claude, ChatGPT, Gemini, specialized tools like Elicit or Consensus) or keep it generic as "LLMs" / "generative AI"? Generic is more durable; specific is more concrete. I'd default to generic in the summaries and selectively specific in the appendices where it clarifies a use-case.

One suggestion

Given this is meant to circulate, consider letting me produce it as a **Word document** at the end so faculty can comment/track-changes on it rather than re-formatting from chat. I can deliver both an inline version (so you can read it immediately) and a .docx (for circulation). Just say the word.

A small flag on the brief itself

The constraint that future capabilities will be "at least as good as Opus 4.7 Extended" is useful but carries a subtle risk worth naming somewhere in the document: capability alone doesn't determine educational impact — *how it's used and how available it is* matter at least as much. A tool that's 10× more capable but used identically to today's tools has a different effect than a tool of equal capability that becomes ambient, always-on, and free. I'll build this into the career-horizon discussion rather than dwell on it, but wanted to flag that I'm reading the constraint as "assume capability floor ≈ current frontier" rather than "assume use patterns stay fixed."

Give me your answers on the four questions (or tell me to use my defaults) and I'll draft.

Last updated on April 29, 2026

1. *Three Appendices. Be expansive. 2. Be expansive in your use of citations. Engaged safeguards to sidestep potential fabrication or hallucinations. 3. Regarding Appendix 3. I would like you to do "both." I would like you to sketch out pro's, con's, and potential prescriptions. 4. Regarding "tool-naming." I agree with your suggestion. 5. Yes, both inline and Word docx. 6. I agree with your reasoning and approach.*

Generative AI in Doctoral Training in Neuroscience and Psychological Science

A Balanced Briefing for Department and Graduate School Faculty

Claude Opus 4.7X & AJ Shackman

April 22, 2026

Executive Summary (Short)

Generative AI is becoming a pervasive feature of scientific training, and neuroscience and psychological science PhD programs must now decide how — not whether — to respond.

The case for integration is strong. Large language models already accelerate coding, drafting, literature triage, and ideation; they democratize technical competencies; and within current students' careers, they will plausibly co-author, critique, and propose experiments. Graduates who finish AI-naïve will enter a workforce expecting fluent, critical human–AI collaboration.

The case for caution is equally serious. Frequent AI use has been linked to diminished critical-thinking performance mediated by cognitive offloading (Gerlich, 2025). EEG evidence suggests that LLM-assisted essay writing produces weaker cortical connectivity and severely impaired recall of one's own output (Kosmyna et al., 2025) — what those authors call *"cognitive debt."* These findings are preliminary and under-replicated, but they converge with older work on transactive memory and a growing literature documenting homogenized outputs, overreliance, integrity risks, and a non-trivial rate of fabricated citations (Linardon et al., 2025). The deepest concern is *hollow fluency*: students who produce polished work they cannot defend or extend.

A balanced posture is neither prohibition nor laissez-faire. We recommend principled integration:

Last updated on April 29, 2026

1. Treat AI literacy as a required doctoral competency.
2. Deliberately protect AI-free contexts — qualifying exams, methods bootcamps, selected writing exercises, oral defenses — to build and verify foundational skills.
3. Establish clear disclosure norms and research-integrity guidance.
4. Invest explicitly in the mentorship and apprenticeship dimensions of training, which LLMs cannot replace.
5. Reassess annually as capabilities and norms evolve.

Because embedding will vary across laboratories and courses, the appendices expand on (1) specific use-cases, (2) sub-disciplinary considerations, and (3) specific foundational skills worth protecting, with prescriptions for each.

Executive Summary (Long)

Framing

Generative AI is rapidly becoming a standard part of how scientific work gets done. Doctoral programs in neuroscience and psychological science are past the point of whether to engage — the question is how to do so in ways that produce rigorous, independent scientists rather than fluent but brittle ones. Because embedding will vary by laboratory, course, and sub-discipline, this briefing offers defaults and principles rather than a one-size-fits-all policy.

Career horizon and assumptions

Students entering PhD programs now will hit their first major career inflection (tenure-track review or its industry equivalent) around 2038–2040. Over that horizon, the minimum reasonable assumption is that frontier LLMs will retain at least the capabilities of current systems (e.g., Claude Opus 4.7 Extended), with ambient availability and deeper integration into laboratory workflows. Capability alone, however, does not determine educational impact — usage norms, institutional defaults, and economic incentives matter at least as much. Recommendations below are therefore robust across plausible capability futures rather than pegged to specific forecasts.

The case for integration

Productivity: LLMs accelerate coding, drafting, literature triage, debugging, and the routine scaffolding of scientific work. **Access:** they democratize technical competencies (statistical

coding, numerical simulation, English-language writing) historically gatekept by expense or prior training, reducing inequities for students from less-resourced backgrounds. **Idea partnership:** used well, they serve as an always-available interlocutor for stress-testing arguments and surfacing adjacent literatures. **Workforce alignment:** fluent, critical human–AI collaboration is plausibly becoming a baseline professional expectation. **Scientific acceleration:** deep-learning methods have already transformed parts of biology — AlphaFold (Jumper et al., 2021) is the canonical example — and are beginning to reshape hypothesis generation, simulation, and analysis in ways students must understand.

The case for caution

Early empirical work is concerning, if not yet definitive. Gerlich (2025, n = 666) found a significant negative correlation between frequent AI use and critical-thinking performance, mediated by cognitive offloading, with effects strongest in younger and more AI-dependent users. Kosmyna et al. (2025), using EEG in a 54-participant study of LLM-assisted essay writing, reported the weakest cortical connectivity in the LLM group and found that 83% of LLM users could not quote passages from essays they had just written. The authors describe this as *cognitive debt*: short-term efficiency purchased at long-term cost in memory, ownership, and autonomous reasoning. These findings are preliminary and methodologically limited (small samples, constrained tasks, specific models), but they converge with older work on transactive memory (e.g., Sparrow et al., 2011 — a classic finding that has not replicated cleanly) and with a growing literature documenting overreliance, output homogenization, and integrity challenges. A separately documented risk is *hallucination*: a recent peer-reviewed audit found ~20% of GPT-4o's citations were entirely fabricated in mental-health literature reviews, with 45% of the remainder containing bibliographic errors and higher fabrication on less-covered topics (Linardon et al., 2025).

Hollow fluency and the viva

The deepest pedagogical worry is harder to measure but widely reported by faculty: students can produce work of increasing surface sophistication while becoming less able to defend, extend, or critique it in real time. For a PhD — where the ultimate product is a trained mind rather than a submitted document — this matters enormously. The viva and conference Q&A are diagnostic precisely because they reveal whether understanding is internalized or merely rented.

The mentorship dimension

LLMs increasingly absorb the "dumb questions" students once took to postdocs and PIs. This has genuine upsides (reduced friction, less embarrassment for struggling students) and real

downsides (erosion of the apprenticeship through which taste, judgment, and tacit knowledge are transmitted). Programs that do not deliberately protect these spaces may produce graduates who are technically competent but lack what distinguishes a scientist from a sophisticated operator.

Epistemic status

The empirical base for strong claims is thin. The best available studies are small, recent, and conducted on tasks that imperfectly model doctoral training. Historical parallels (pocket calculators, Google, Wikipedia) should make us humble about both alarmist and utopian predictions. The honest posture is provisional integration with serious ongoing monitoring.

Relationship to the replication crisis

Both fields, and psychology especially, have been reshaped by the recognition that much published work may not replicate (Open Science Collaboration, 2015; with a methodological rejoinder by Gilbert et al., 2016). AI cuts both ways: it can enable larger pre-registered replications, automated robustness checks, and better documentation — or it can accelerate sloppy pipelines, facilitate p-hacking, and generate fabricated citations. How students are trained to use these tools will materially affect whether AI becomes a force for reform or its opposite.

Recommended posture: principled integration

We suggest programs adopt neither prohibition nor laissez-faire, but defaults that can be adapted across laboratories and sub-disciplines:

6. Treat AI literacy — prompt construction, verification, awareness of failure modes, research-ethics implications — as a required doctoral competency, alongside statistics and scientific writing.
7. Deliberately protect AI-free contexts where foundational skills are built and verified: first-year methods bootcamps, qualifying exams, oral defenses, and selected writing exercises. These are diagnostic and developmental, not punitive.
8. Publish explicit, lab-level disclosure norms and research-integrity guidance. Require documentation of AI use in training and thesis work, and verification of all AI-provided citations against primary sources.
9. Invest explicitly in mentorship and apprenticeship — regular unstructured time with advisors, lab traditions of unaided reasoning, informal journal clubs — that LLMs cannot replicate.

10. Teach students to use LLMs for productive struggle, not bypass: attempting tasks first, then comparing with AI output; using AI as a critic of their work rather than a producer of it.
11. Reassess annually. Capabilities, usage norms, and the empirical literature are all moving faster than curricula typically do.

The three appendices that follow expand on specific use-cases, sub-disciplinary considerations (recognizing the internal heterogeneity of both fields), and the specific foundational skills most worth protecting — with prescriptive guidance for each.

Appendix 1: Pros and Cons by Specific Use-Case or Scenario

This appendix surveys common ways LLMs are already being embedded in doctoral training. Each scenario carries a distinct risk/benefit profile; laboratories should be explicit about which they permit, require, or discourage. In each entry below, italics indicate field labels (Pros, Cons, Note) rather than direct quotations.

1. Permitted use with disclosure (low-touch policy)

Pros: Reflects the likely workplace reality for graduates; avoids forcing students into hidden, unsupervised use; preserves autonomy.

Cons: Without scaffolding, students default to the most cognitively offloading uses (production over critique); disclosure norms drift without enforcement.

Note: Works best when paired with explicit AI-literacy training and protected AI-free contexts.

2. Coding and scripting co-pilot

Pros: Large productivity gains for routine scripting (data cleaning, plotting, standard statistical analyses); flattens the learning curve for students new to Python, R, or MATLAB; reduces time lost to syntax.

Cons: Students may never build a bottom-up mental model of what their code does; silent errors and “plausibly-wrong” analyses become easier to produce at scale; debugging skills atrophy.

Note: Consider a “first-pass unaided” norm for coursework, followed by AI-assisted refactoring for comparison and critique.

3. Writing assistance (manuscripts, grants, dissertation chapters)

Pros: Lowers barriers for non-native English speakers (a significant equity gain); accelerates drafting; provides useful critique at low cost.

Cons: Homogenizes voice; risks authorial detachment from the argument; Kosmyna et al. (2025) suggests heavy LLM use in writing may impair recall and ownership of one's own text.

Note: Distinguish AI as editor/critic (generally low-risk) from AI as first-draft generator (higher risk, especially at developmental stages).

4. Literature review and discovery

Pros: Rapidly surfaces adjacent fields, summarizes unfamiliar methods, identifies connections a narrow search would miss.

Cons: LLMs fabricate citations at non-trivial rates — Linardon et al. (2025) found ~20% outright fabrication and ~45% additional bibliographic errors in GPT-4o-generated mental-health literature reviews, with higher fabrication on specialized/less-covered topics; misattribution of findings is even more common than outright fabrication.

Note: Every AI-surfaced reference must be verified at primary source before use. Consider requiring students to maintain a dual bibliography: cited works and AI-surfaced leads they have independently verified.

5. Hypothesis generation and experimental design

Pros: LLMs can generate large numbers of plausible hypotheses; stress-test designs against common pitfalls; surface alternative explanations. Used as a critic ("what's wrong with this design?"), they substantially reduce undergraduate-level errors.

Cons: Generated hypotheses are often superficially novel but epistemically derivative, pulling toward the training distribution's center of mass; reliance may narrow the range of ideas students generate independently.

Note: Encourage use of LLMs as devil's-advocate critic after a student has independently formulated a hypothesis — not as a generator of first ideas.

6. Tutoring and teaching-assistant role

Pros: 24/7 patient explanation at adjustable depth; willingness to engage with "dumb questions" students might hesitate to ask a PI; can adapt examples to individual learning styles.

Cons: Displaces precisely the mentorship interactions through which taste and judgment are transmitted; may reinforce superficial understanding when students mistake the AI's fluent explanation for their own comprehension.

Note: Pair AI tutoring with periodic unaided problem sets to calibrate students' actual understanding.

7. Data analysis and statistical guidance

Pros: Democratizes access to sophisticated methods (Bayesian modeling, mixed-effects, multivariate analyses) historically confined to statistically trained labs; can flag assumption violations students might miss.

Cons: Enables analyses the student cannot justify or audit; high risk for p-hacking and "pipeline sprawl"; silent errors in cross-tool translation.

Note: Require a "methodological audit trail" for any AI-assisted analysis, including rationale for each analytic choice and confirmation the student can reproduce key steps unaided.

8. AI as reasoning partner or critic

Pros: Potentially the highest-value use; a tireless interlocutor for refining arguments, surfacing counter-examples, pressure-testing logic. Approximates a senior lab-mate who will always read your draft.

Cons: Effectiveness depends on the student's existing critical capacity — students who lack foundational skills cannot meaningfully evaluate or push back on AI output, producing an illusion of rigor.

Note: This mode rewards students who have already built strong independent skills, reinforcing the case for protected AI-free developmental contexts early in training.

Appendix 2: Sub-Disciplinary Considerations

Neuroscience and psychological science are internally heterogeneous; blanket policies will fit some sub-fields poorly. This appendix sketches differential considerations. Italics below indicate the recommended training emphasis for each sub-field.

Last updated on April 29, 2026

Systems and circuit neuroscience (animal models, wet-lab)

The hands-on craft of surgery, histology, electrophysiology, and behavior remains substantially physical and protocol-bound — less immediately displaceable by LLMs. However, computational analysis pipelines (spike sorting, behavioral scoring, connectomics) are rapidly being transformed, and animal-welfare, regulatory, and protocol-writing components benefit from AI assistance.

Training emphasis: Preserve wet-lab apprenticeship; integrate AI in analysis and writing with standard guardrails.

Human neuroimaging (fMRI, EEG/MEG, fNIRS)

Heavily computational; large upside for AI assistance in preprocessing, modeling, and visualization. Also the area most vulnerable to automated bad practice — multiple-comparison mishandling, pipeline sprawl, silent reproducibility failures.

Training emphasis: Rigorous statistical foundations first; pre-registration and computational reproducibility norms; AI as accelerator only after these are established.

Theoretical and computational neuroscience

Closest to being fundamentally restructured. LLMs are already capable collaborators for modeling, derivation, and code, and this will only deepen.

Training emphasis: First-principles mathematical training (dynamical systems, information theory, machine-learning theory) remains essential and should be defended vigorously; AI accelerates application but cannot substitute for the conceptual base.

Cognitive psychology and cognitive neuroscience (behavioral, EEG)

Intermediate position. Experimental design, paradigm construction, and careful measurement remain core skills. AI assists in stimulus generation, piloting, and analysis.

Training emphasis: Protect the hand-crafting of paradigms as a skill; students should be able to run a study end-to-end without AI at least once.

Social and personality psychology

Intersects intensely with the replication-crisis legacy. AI can help (pre-registration tooling, automated robustness checks, larger-scale replications) or hurt (plausible-sounding but invalid theoretical framings; fabricated citations in introductions and discussions).

Training emphasis: Statistical rigor, theoretical depth, and source verification as non-negotiable competencies.

Developmental psychology

Much of the field requires direct engagement with children and families — difficult to displace. AI aids coding of behavioral data, longitudinal analysis, and literature synthesis.

Training emphasis: Standard integration with particular attention to measurement validity, which LLMs handle poorly.

Clinical psychology and clinical science

Additional layers: patient-data privacy (HIPAA and equivalents), therapeutic training that is inherently relational, and regulatory constraints on AI-assisted clinical decisions. AI use in clinical training raises issues that do not arise in basic science.

Training emphasis: Explicit clinical-AI ethics curriculum; strict guidance on protected health information and AI tools; preservation of supervised clinical hours as AI-free.

Quantitative psychology and psychometrics

Historically a small, specialized area; LLMs lower the barrier for non-specialists to engage with advanced methods, which is both a pro (wider adoption of good methods) and con (wider adoption of methods that are misapplied).

Training emphasis: Strengthen the quantitative core so these students remain the methodological conscience of their departments.

Appendix 3: Foundational Skills — Threats, Tradeoffs, and Prescriptions

Not all "foundational skills" are equally threatened or equally worth protecting. This appendix inventories specific skills, estimates their exposure to AI-driven erosion, and offers prescriptive guidance. The framing throughout: the goal is not to freeze the skill set of 2019 but to decide deliberately which capacities are constitutive of being a scientist and must be actively defended.

1. Statistical intuition built through hand calculation

Threat level: High. Students can now complete most coursework without ever manipulating an estimator by hand.

Why it matters: Intuition about what a test is doing, when assumptions matter, and why a result is surprising comes from working through the math, not from reading output. Without it, students cannot audit AI-assisted analyses or catch silent errors.

Prescription: Maintain AI-free problem sets in first-year methods courses; require derivation of core estimators (OLS, basic hierarchical models, permutation tests) by hand; include hand-computed sanity checks in written qualifying exams.

2. First-principles derivation

Threat level: High for computational and theoretical sub-fields; moderate elsewhere.

Why it matters: The ability to derive a result from assumptions is what separates scientists from operators. It is also what allows recognition of when an AI output is wrong.

Prescription: Protect chalk-and-talk seminars; require derivation-based components in qualifying exams; explicitly teach students to attempt derivations before consulting AI.

3. Close reading of primary literature

Threat level: High. LLM summaries are seductive substitutes for actual reading.

Why it matters: Theoretical taste, methodological judgment, and the ability to notice what a paper does not say are all built through close reading. Summaries homogenize and flatten.

Prescription: Journal clubs with explicit no-AI-summary rules; “cold reading” exercises where students discuss a paper they have read without AI assistance; advisor practices of discussing papers in the raw.

4. Literature synthesis across sources

Threat level: Moderate–high. AI is genuinely good at surface synthesis; the distinctive human skill is judgment about weight, quality, and contested claims.

Why it matters: This is the heart of the qualifying exam and the introduction to a dissertation. Offloaded, it becomes a veneer.

Prescription: Require synthesis exercises with explicit argumentation that contradicts or complicates a literature; maintain oral defense of comprehensive exams.

5. Experimental design and critique

Threat level: Moderate. AI can critique well, but students who lean on it early may never develop independent design judgment.

Why it matters: This is the most distinctive doctoral-level skill, and the one most employers care about in PhD graduates.

Prescription: First-year design exercises where students propose, critique, and revise designs without AI; AI as second-pass critic only; oral defense of first-year project proposals as diagnostic.

6. Coding — syntax recall vs. conceptual understanding

Threat level: Low for the syntax-recall component (and arguably not worth defending); high for the understanding-from-first-principles component.

Why it matters: Looking up syntax was always semi-outsourced. Understanding what one's analysis pipeline actually does is not.

Prescription: Emphasize code-review and reproducibility as core skills over syntax fluency; require students to walk through their pipelines line-by-line in oral examinations.

7. Scientific writing

Threat level: Mixed. AI as editor/polisher is low-risk and genuinely equitable. AI as first-draft generator is high-risk — the act of drafting is often where thinking happens.

Why it matters: "I don't know what I think until I try to write it" is not a cliché; it is a cognitive fact consistent with Kosmyna et al.'s (2025) observations.

Prescription: Protect certain writing exercises — first drafts of grant specific aims, abstracts of one's own work, initial dissertation chapters — as AI-free. AI-assisted polishing after the thinking is done is fine.

8. Oral defense and thinking on one's feet

Threat level: High, indirectly. If students can produce AI-polished work they have not fully internalized, the gap shows in oral examination.

Why it matters: This is where hollow fluency is most diagnostic.

Prescription: Preserve and strengthen oral components throughout training — qualifying exams, lab meetings, conference practice. Consider a brief oral "internalization check" as part of dissertation proposal review.

9. Productive struggle and tolerance for confusion

Threat level: High. Instant AI help short-circuits the productive discomfort through which understanding is built.

Why it matters: Much of what we recognize as deep expertise is the residue of having sat with confusion long enough to restructure one's mental model. LLMs can inadvertently remove this developmental pressure.

Prescription: Build explicit "struggle windows" into training — time-boxed intervals (e.g., 30–60 minutes) during which students attempt a problem unaided before consulting AI or peers. Normalize and model this practice at the faculty level.

10. Mentorship and apprenticeship transmission

Threat level: High, through a different mechanism. LLMs do not erode the skill directly; they erode the occasions on which the skill is transmitted, by absorbing the questions and uncertainties that used to bring students to senior colleagues.

Why it matters: Taste, judgment, ethical sensibility, and the tacit craft of science have always been passed down through sustained one-on-one contact. The cost of losing this is invisible in the short run and catastrophic in the long run.

Prescription: Deliberately protect low-stakes, unstructured interactions between students and advisors (walking meetings, lab-meeting whiteboarding, standing office-hours norms). Counter-intuitively, one of the strongest pro-mentorship moves a program can make in the AI era is to create more — not fewer — occasions for face-to-face reasoning that has no AI-mediated alternative. Faculty should model AI use AND its deliberate limits in front of students.

References

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science." Science, 351(6277), 1037. https://doi.org/10.1126/science.aad7243

Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv:2506.08872.

Linardon, J., Jarman, H. K., McClure, Z., Anderson, C., Liu, C., & Messer, M. (2025). Influence of Topic Familiarity and Prompt Specificity on Citation Fabrication in Mental Health Research Using Large Language Models: Experimental Study. JMIR Mental Health, 12, e80371. https://doi.org/10.2196/80371

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. Science, 333(6043), 776–778. https://doi.org/10.1126/science.1207745