Journal of Technology-Integrated Lessons and Teaching
(2026) Volume 5, Issue 1; DOI: 10.13001/jtilt.v5i1.10297

Teaching Students to Identify Ethical Risks and Blind Spots in Academic AI Use

Sara G. McNeil, University of Houston

Overview

This lesson equips graduate students across disciplines to navigate ethical challenges when using generative AI tools in their research and writing. Through guided discussion, students learn five "AI Personas" (e.g., Research Assistant, Writing Coach, Data Detective, Idea Generator, Academic Concierge) representing common AI use cases. For each persona, students identify specific ethical blind spots and develop mitigation strategies. The technology-rich experience centers on real-time polling, enabling anonymous sharing of concerns and collaborative scenario analysis. Students receive an *AI Ethics Checklist* and an *AI Assistance Log Template*. The assessment includes a pre-/post-poll comparison to demonstrate any changes in ethical reasoning.

Topics: AI ethics, responsible AI use, documentation practices for AI, graduate student use of AI in research and writing

Time: 30 minutes for core content, 45-60 minutes with extended activities.

Materials

- Computer with projector and screen
- Internet connection
- Polling software (Poll Everywhere or similar)
- Student devices
- Whiteboard or flip chart with markers
- Presentation slides
- AI Personas Quick Reference Guide
- Daily AI Ethics Checklist
- AI Assistance Log Template
- Scenario Cards for Discussion

Context-at-a-Glance

Setting
Workshop for doctoral students at a large public research university

Modality
In-person delivery

Class Structure
30-minute session during a monthly doctoral student professional development series

Organizational Norms
The university emphasizes research integrity and innovation and provides professional development support for graduate students. Faculty encourage exploration of emerging technologies while maintaining academic standards.

Learner Characteristics
50-75 doctoral students in education (EdD), diverse backgrounds, varying AI experience from none to daily use.

Instructor Characteristics
Instructor has 20+ years teaching graduate research methods, extensive experience mentoring doctoral students (50+ dissertations), and recent engagement with AI ethics in academic contexts.

Development Rationale
Created to address urgent student questions about AI ethics in academic use amid the absence of clear institutional policies. The training helps students develop judgment-based decision-making frameworks.

Design Framework
Situated learning theory combined with metacognitive reflection strategies, using concrete personas to ground abstract ethical principles in recognizable research contexts.

Setup

Environment Setup: Arrange seating so that all students can see the projection screen clearly and comfortably. Ensure adequate spacing for note-taking and occasional pair discussions. The room should have controllable lighting (bright enough for device use, dim enough for screen visibility) and minimal audio distractions. Post polling access information (URL and text code) prominently where students can reference it throughout the session.

Technical Setup: Load presentation slides and test all transitions. Log in to the polling platform and load all poll questions (pre-assessment, scenario question, post-assessment) or integrate polling in the presentation. Test poll accessibility from a mobile device to ensure students can connect. Verify internet connectivity is sufficient for simultaneous polling connections. Have a backup plan ready (e.g., Google Form, hand-raising, paper cards) in case of technical failure.

Time Estimate: Allow 15-20 minutes for complete setup. This includes room arrangement (5 minutes), technical testing (10 minutes), and material preparation (5 minutes). Instructors teaching this lesson for the first time should allow an additional 10 minutes to familiarize themselves with the polling platform features and slide presenter notes.

Context and Setting

This lesson was developed and implemented at the University of Houston, a large public research university (R1 institution) serving a diverse metropolitan area. The College of Education houses both EdD (Doctor of Education) and PhD (Doctor of Philosophy) programs, with approximately 150 active doctoral students at any time. The institutional culture emphasizes rigorous scholarship and practical application, reflecting Houston's role as a major urban center with significant educational needs.

When this lesson was developed, the university had not established formal policies on AI use in graduate education. This policy vacuum created anxiety among students who were simultaneously hearing messages that AI was "the future of research" and warnings about academic integrity violations. Faculty were equally uncertain, with some enthusiastically embracing AI tools while others banned them entirely.

The target audience was doctoral students in education, although the lesson has been adapted for students at other levels and in other disciplines. These students ranged from recent master's graduates in their late 20s to mid-career professionals in their 50s pursuing doctoral degrees while working full-time. Many were practicing educators, administrators, or instructional designers, balancing dissertation work with demanding professional roles.

Most students previously encountered generative AI tools like ChatGPT, Claude, or Gemini, but their experience ranged dramatically. Some had never used AI beyond asking simple questions, while others incorporated AI into their daily research workflow. However, even experienced AI users expressed uncertainty about ethical boundaries. Common questions included: "Can I use AI to summarize articles?" "Is it cheating if AI helps me reorganize my chapter?" "What do I need to tell my advisor?" "How do I cite AI assistance?"

Students' anxiety about AI ethics was compounded by the high-stakes nature of doctoral work. Unlike undergraduate assignments, which can be revised or retaken, dissertation chapters represent years of work and a significant financial investment. The fear of making an irreversible ethical mistake, one that could jeopardize degree completion, created paralysis for some students who avoided AI entirely, even when it might legitimately support their learning.

Delivery Format and Constraints

The lesson was designed for the College's "ADVANCE" professional development meeting, which brings doctoral students together for skills training beyond their regular coursework. These sessions are voluntary and compete with students' research, teaching, and work obligations. The time constraint was the most significant design challenge since I wanted to address basic ethical questions in 30 minutes while maintaining engagement and providing practical takeaways for students.

The physical setting was a large classroom with stadium seating, a large projection screen, and reliable Wi-Fi supporting 100+ simultaneous connections. Students brought their own devices

(laptops, tablets, or smartphones) and were comfortable with technology-mediated participation. The informal, supportive atmosphere of the ADVANCE series allowed for honest discussion of challenges and uncertainties without fear of judgment.

Instructor Positioning

As the instructor, my 20+ years of experience mentoring doctoral students through dissertation completion provided credibility when discussing ethical dilemmas. Students trusted that I understand the pressures they face, the relationships with advisors they must maintain, and the high stakes of their work. However, I also positioned myself as learning alongside them. I shared my own uncertainties about AI ethics and acknowledged that best practices are still emerging. I believe this stance reduced their defensiveness and created space for exploration rather than compliance with prescribed rules.

My moderate (but not expert) technical proficiency with AI tools proved advantageous. Students needed guidance from someone who understood both the technology and the academic context, but not from someone so technically advanced that recommendations felt unattainable. I felt I could model thoughtful AI use without intimidating students who were just beginning to explore these tools.

Design Decisions

The design framework combined situated learning theory, which grounds learning in authentic contexts and real problems, with metacognitive reflection strategies that prompt learners to monitor and evaluate their own thinking and decision-making. Together, this framework structured students' thinking through concrete personas while incorporating deliberate pauses for self-assessment.

That framework guided my decisions as I developed the lesson. Five key decisions reflect how I translated situated learning and metacognitive reflection into a practical and engaging session for doctoral students.

Personas Framework

Rather than abstract ethical principles, I developed five "AI personas" that represent specific use cases students would encounter in their research. This helped to address the disconnect between theory and practice. Students could immediately identify which persona matched their situation and use the blind spots and practical responses to think through their own AI use.

Interactive Polling

Opening with a poll allowed students to share their concerns anonymously and see that their peers held a wide range of perspectives. The real-time display created a sense of community around what might otherwise have felt like an isolating uncertainty. Doctoral students often assume everyone else has figured out the rules while they are still struggling with the boundaries. Seeing responses range from genuine anxiety to enthusiastic embrace helped normalize the conversation before it began. This sense of community and the anonymity of the responses provided psychological safety for students to engage openly and honestly throughout the session.

Judgment Over Rules

I designed this lesson to cultivate judgment rather than provide rules, since students needed frameworks that would work regardless of specific context or evolving AI capabilities. The four "red flag questions" reflect that approach:

- If my advisor asked me to explain this right now, could I?
- Would I be comfortable defending this choice in my defense?
- Am I using AI to learn faster or to avoid learning?
- Could someone reproduce my work based on my documentation?

These questions functioned as metacognitive prompts: they required students to step back from task completion to assess their own understanding, honesty, and process.

Documentation Emphasis

I emphasized documentation because it works across disciplines and advisor expectations. In the absence of clear institutional policies, students needed a personal record of their AI use that could withstand scrutiny from their advisor or committee.

The AI Assistance Log template provides that structure, making documentation feel manageable rather than burdensome, making their processes transparent to advisors, and building habits of accountability. The log also serves as a metacognitive tool, prompting students to reflect on their purpose, process, and learning each time they use AI.

Practical Takeaways

I developed the Daily AI Ethics Checklist and documentation template to provide students with immediately usable tools that justify their time investment in attending. The Checklist extends the reflective practice from the session into students' ongoing work, prompting them to evaluate purpose, attribution, understanding, and growth each time they use AI.

Together, these five decisions reflect the situated learning framework's emphasis on authentic, contextually grounded practice, using real student concerns and realistic scenarios to make abstract ethical principles concrete. The metacognitive reflection strategies are built into each phase: The opening poll reveals students' existing assumptions, the red flag questions interrupt task-focused thinking and prompt self-assessment, and the Daily AI Ethics Checklist sustains that reflective practice beyond the session itself.

Learning Representation

Phase 1: Introduction (3-4 minutes)

Learning Objective: Students will explore their existing concerns about AI use and establish a baseline understanding of ethical considerations.

Activity: Opening Poll

I opened the session using Poll Everywhere (https://www.polleverywhere.com/), displaying the polling slide with a web URL and text message access option to accommodate different devices. This tool worked well for this purpose, though the free version limited responses to 40 per session.

I asked students: *In a few words, describe your biggest concern about using AI in your doctoral work.*

As responses accumulated over one to two minutes, I commented on them in real time, noting emerging themes aloud to build engagement and signal that their concerns were heard. After collecting responses, I identified two or three common themes using students' language and explicitly connected them to the session ahead. I could have captured the themes on a whiteboard for visual reference during the Phase 3 content presentation.

The response patterns served as a baseline indicator of students' sophistication with AI ethics. Surface-level concerns ("getting in trouble") suggested a need for foundational framework-building, while metacognitive concerns ("losing my voice") indicated students were already thinking about AI's impact on their scholarly development. I used this read to shape my emphasis during the personas discussion that followed.

Transition

I closed the opening by asking students to keep their concerns in mind throughout the session, and I framed the personas framework ahead of time as a tool for answering many of those questions for themselves.

Phase 2: Framing the Challenge (1-2 minutes)

Learning Objective: Students will establish a pragmatic, non-judgmental stance toward AI use and clarify the session focus.

Content Presentation

I displayed a slide with news headlines that presented conflicting narratives: AI as an existential

threat and AI as a transformative opportunity. A "Reality Check" slide followed, anchoring the session's central message that AI is not going away, and the relevant question is not whether to use AI in doctoral work but how to use it ethically and effectively.

Then I displayed the learning objective, framing the session explicitly as being about developing judgment rather than following rules, neither avoiding AI nor embracing it uncritically. This framing proved important for reducing defensiveness and creating space for honest engagement with the personas that followed.

Phase 3: Core Content - The Five AI Personas (15-18 minutes)

Learning Objective: Students identify specific use cases, ethical risks, and mitigation strategies for five common AI applications in research.

Instructional Structure

For each persona, I followed a consistent three-minute pattern: briefly describing what the AI tool does in that role, acknowledging why it is genuinely tempting for doctoral students, identifying the ethical blind spots students most commonly miss, and suggesting practical ways to address those risks. This predictable structure helped students take effective notes and compare across personas as we moved through them.

Persona 1: The Research Assistant (3 minutes)

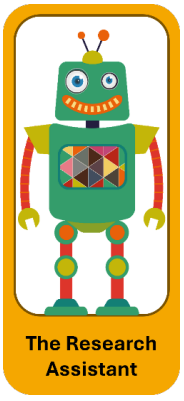

Figure 1. The Research Assistant.

I introduced the Research Assistant persona (see Figure 1) as the one most students had already encountered. Artificial intelligence is used to summarize articles, identify themes across literature, suggest search terms, and make an overwhelming body of literature feel manageable. When I asked for a show of hands on who had already used AI for these purposes, most students responded affirmatively. This helped normalize the conversation.

The temptation with this persona is worth acknowledging: What used to take weeks to complete can feel like it takes hours with AI. But I focused on three blind spots that users may miss if they use AI to summarize multiple sources:

The first is what I called the Synthesis Trap, consuming AI's interpretation of the field rather than personally wrestling with the ideas. Students might summarize sources fluently but struggle when others push them to engage in genuine scholarly debate.

The second is citation integrity: when AI synthesizes fifty papers, it can be unclear whose ideas are being built upon and whether crucial methodological details were considered.

The third is the knowledge gaps and biases that result. AI systems reflect biases in their training data, which means students risk framing their research around those biases rather than interrogating them. A study's 30% response rate or convenience sample may not be included in an AI summary, and entire perspectives or bodies of scholarship may be underrepresented or missing entirely. Reading against AI-generated themes rather than simply seeking confirmation is an important check.

I offered students four practical suggestions:

1. Treat AI summaries as starting points rather than endpoints,
2. Always go back to original sources before citing them,
3. Test your own understanding by asking whether you could explain the concept to your advisor without the summary, and
4. Document your process so your reasoning is transparent, including any steps you took to identify and check against potential AI bias in the themes that emerged.

Persona 2: The Writing Coach (3 minutes)

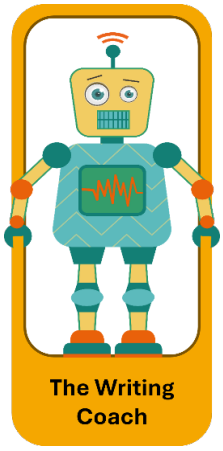

Figure 2. The Writing Coach.

The Writing Coach (see Figure 2) is the AI persona that improves sentence structure, suggests transitions, helps with academic tone, and breaks through writer's block at 2 a.m. Students connected with this persona immediately. The temptation is obvious: AI fixes awkward phrases instantly and writing sounds more polished and academic. But I focused on three blind spots that are easy to miss.

The first is voice authenticity. If AI reorganizes your arguments, improves your word choice, and creates your transitions, at what point does it become AI's writing rather than your own?

The second is intellectual honesty. There is a difference between learning to write academically and learning to prompt AI to write for you. Your advisor expects to see your thinking process, including the messy parts, because that is how they know where you are struggling and how you are developing.

The third is committee expectations. Your committee is not just evaluating your conclusions; they are evaluating how you think. When they read only perfect, polished prose, they lose visibility into your reasoning process, and you lose the feedback that helps you develop as a scholar. Sometimes a messier draft reveals more about your thinking than AI's cleaned-up version.

I offered four practical suggestions:

1. Use AI for mechanics like grammar and transitions,
2. Save pre-AI versions and ask yourself whether you can still recognize your own thinking,
3. Be transparent with your advisor about what assistance you are using
4. Keep an AI assistance log. The log is more than a record because it prompts the kind of metacognitive reflection the red flag questions are designed to encourage. Before accepting AI's changes, students should be able to answer: Could I explain this reorganization to my advisor right now? Would I feel comfortable defending it in my defense? Those questions turn the Writing Coach from a shortcut into a learning tool.

Persona 3: The Data Detective (3 minutes)

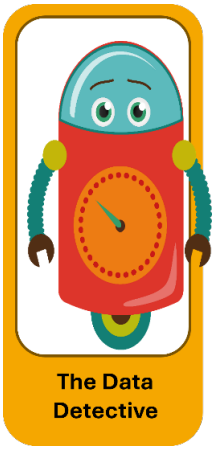

Figure 3. The Data Detective.

The Data Detective persona (see Figure 3) suggests statistical approaches, identifies patterns in datasets, explains complex analyses, generates analysis code, and catches errors. I noted that this persona is particularly relevant for quantitative work, but increasingly valuable for qualitative methods as well. I focused on three blind spots.

The first is what I called the Black Box Problem. If AI suggests an analytical approach and you cannot explain why you chose it, you may be in trouble when your committee asks you methodological questions regarding your work. "Because AI suggested it" is not a defense!

The second is methodological rigor. AI tends to suggest approaches that are good enough rather

than most appropriate for your theoretical framework and your discipline's standards. It does not know your field the way your methodologist does.

The third is reproducibility. If your analysis depends on AI-generated code that you do not fully understand, it may be difficult for others to replicate your work, and replicability is a cornerstone of research integrity.

I offered four practical suggestions:

1. Never use a method you cannot explain,
2. Verify AI suggestions against disciplinary best practices and your committee's guidance,
3. Document your prompts and AI's reasoning alongside your own verification steps,
4. Test yourself by asking whether you can explain why the method is appropriate, interpret the results, and defend the choice.

The goal is for AI to introduce you to methods you then make your own through study and consultation, not for AI to make methodological decisions for you. There is a meaningful difference between saying "AI told me to use this approach" and saying "AI introduced me to this method, I studied it, consulted with my advisor, and chose this approach because it fits my theoretical framework and research questions." A developing scholar should work by using available tools to accelerate learning while maintaining ownership of every methodological decision.

Persona 4: The Idea Generator (3 minutes)

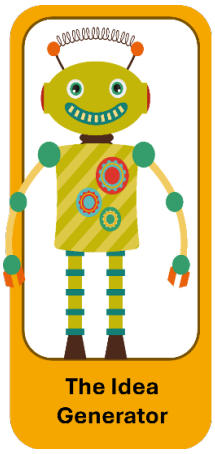

Figure 4. The Idea Generator.

The Idea Generator persona (see Figure 4) brainstorms research questions, suggests theoretical frameworks, proposes novel connections, and breaks through creative blocks. I described this persona as a brainstorming partner who is available at 2 a.m., can offer perspectives from outside your field, and can generate dozens of options quickly. That availability is genuinely appealing to doctoral students who often feel isolated in their thinking. I then focused on three potential blind spots.

The first is originality. When you brainstorm with a human colleague, ideas emerge from genuine intellectual exchange. AI brainstorming works differently because AI tools cannot contribute original thought the way a human collaborator can. If AI tools suggest a theoretical framework and you develop it extensively, questions about whose insight it represents become complicated.

The second is feasibility. AI tools do not know your actual resources, timeline, or committee's expectations. They may suggest ideas that are completely impractical for a doctoral dissertation (e.g., a five-year longitudinal study, access to populations you cannot reach, methods requiring expertise you do not have time to develop).

The third is research currency. AI's training data does not know your subfield's literature the way active scholars do. An idea that seems novel may have already been explored, contested, or abandoned for reasons that only emerge through deep engagement with current scholarship. Before investing time in an AI-suggested direction, complete a thorough literature search. I offered four practical suggestions:

1. Treat AI-generated ideas as raw material that requires your disciplinary expertise and critical analysis,
2. Reality-test ideas with your advisor and peers who know your actual constraints,
3. Conduct thorough literature searches before committing to an idea,
4. Document the origin of ideas clearly so you can articulate your own intellectual contribution.

Persona 5: The Academic Concierge (2-3 minutes)

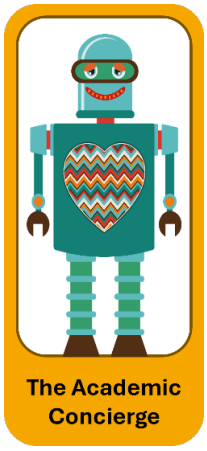

Figure 5. The Academic Concierge.

I introduced the Academic Concierge persona (see Figure 5) as the most straightforward persona and noted that it could be abbreviated if time was running short. This persona uses AI as an administrative assistant: formatting citations, checking grammar and style, managing administrative writing, and handling tedious tasks that consume mental energy. I focused on three blind spots.

The first is skills atrophy. Academic competencies such as citation formatting and style conventions are skills you may need independently throughout your career (in contexts where AI tools are not available or appropriate). Faculty positions, editorial roles, and professional writing all assume a baseline mastery of these skills. Gaps may become visible at bad moments.

The second is quality control. AI tools make mistakes, specifically with citations: AI tools may invent digital object identifiers (DOIs), get dates wrong, and misattribute authors. I shared a concrete example here: A student whose dissertation contained a single fabricated citation generated by AI. That single error delayed the defense by three months while every citation was verified.

The third is professional development. Learning to cite different sources and to follow academic conventions, such as APA, is part of doctoral training. Understanding why citations work the way they do makes individuals more careful and credible scholars. It also builds the kind of attention to detail that carries over into every aspect of research, from data management to peer review. I offered four practical suggestions:

1. Maintain competency by doing some tasks manually on a regular basis,
2. Always verify citations against original sources,
3. Follow what I called a 70/30 rule where you handle the majority of routine tasks yourself
4. Use AI tools selectively, and redirect the time AI tools saves you toward higher-level skill development—rather than just efficiency.

Phase 4: Application and Integration (8-10 minutes)

Learning Objective: Students apply the personas framework to realistic scenarios and receive practical decision-making tools.

Activity 1: Universal Red Flags (2 minutes)

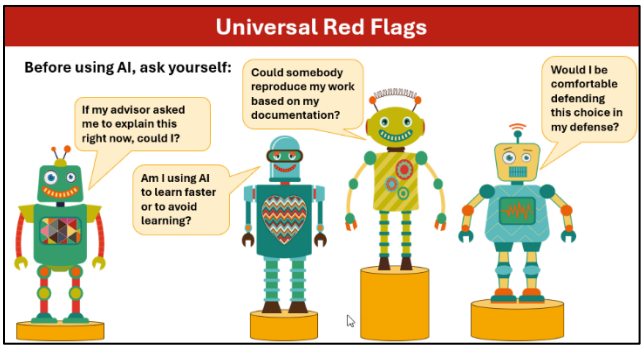

Figure 6. Four universal red flags.

I transitioned to a set of four questions I called universal red flags, displayed on a slide showing each of the five persona robots posing a question (see Figure 6). I framed these as applying to any AI use, regardless of which persona is involved.

The first question is: If my advisor asked me to explain this right now, could I? Not tomorrow after you study, but right now. If the answer is "no," that is a signal to stop using AI and start learning.

The second is: Would I be comfortable defending this choice in my defense? I encouraged students to sit with their emotional response to that question, because discomfort is itself useful information.

The third is: Am I using AI to learn faster or to avoid learning? I told students this is the most important question of the four because one path develops you as a scholar, while the other quietly undermines you. It requires honesty with yourself.

The fourth is: Could someone reproduce my work based on my documentation? If your process depends on AI interactions you have not recorded, your work is not reproducible, which is a fundamental problem regardless of discipline. I framed these four questions as an ethical compass rather than a checklist, something to return to before any AI interaction, not after the fact.

Activity 2: Interactive Scenario Analysis (3-4 minutes)

I presented a scenario through the polling platform to give students a chance to apply the framework to a realistic situation. The scenario was:

Maria uses AI to restructure her literature review chapter. The AI reorganizes her arguments, improves transitions, and fixes awkward sentences. The ideas and evidence are still hers. Does Maria need to acknowledge this AI assistance, and if so, how?

Student responses fell into predictable patterns: some said yes, in a footnote; some said no, because it was just editing; others said to tell the advisor but not document it in the dissertation itself. Rather than immediately providing an answer, I used a think-aloud to model how to apply the framework.

Which persona is Maria using? The Writing Coach.

What are the blind spots? Voice authenticity and intellectual honesty.

Applying the red flag questions:
Could Maria explain the reorganized arguments to her advisor? Probably yes, since the ideas are hers. But did the reorganization reveal a logical structure she was not seeing herself? That is where it gets complicated.

I was direct with students that there is no universal right answer to Maria's situation. It depends on her program's policies, her advisor's expectations, and crucially, what she learned from the process. If AI helped her see weaknesses in her organization that she then understood and could explain, that is learning. If she accepted the changes without understanding why the new structure worked better, that is a problem.

For documentation, I suggested that Maria record the assistance in her process notes and inform her advisor, even if it does not appear in the dissertation itself. The point of the scenario was not to resolve it definitively but to demonstrate that the framework gives students a principled way to think it through, and that developing that habit of reasoning matters more than arriving at any single right answer.

Activity 3: Daily AI Ethics Checklist (2 minutes)

Figure 7. The daily AI ethics checklist.

I distributed the Daily AI Ethics Checklist (see Figure 7), either printed or as a digital download, and walked students through its four prompts. I asked students to be specific about purpose before using AI, thinking not just "help with literature review" but "summarize three articles to identify themes I will look for when I read them fully." I asked them to think through attribution before they begin, not after they used AI. I asked them to consider whether they would still understand their own work after AI assistance. I also asked them to reflect honestly on whether they were learning or just getting things done.

Activity 4: Key Takeaways (2 minutes)

I closed the application phase by presenting four takeaways on a slide. Each AI persona has specific

benefits and blind spots, so there is no universal answer about whether AI is good or bad in doctoral work.

The biggest risks are losing understanding, authenticity, and professional development. Ethical AI use is possible but requires intentionality and documentation. Finally, the goal is for AI to enhance your thinking, not replace it.

I gave students thirty seconds to reflect on which takeaway resonated most and which one would be hardest to apply in their current work.

Phase 5: Closing and Next Steps (2-3 minutes)

Learning Objective: Students leave with concrete action steps and a clear path for continuing to develop ethical AI practices.

I closed the session by displaying four action steps on a slide.

I asked students to keep the Daily AI Ethics Checklist somewhere visible, not filed away. I encouraged them to start an AI assistance log immediately, even if their program does not require one, framing it as a two-minute per-entry task that will save hours of uncertainty later. I shared a QR code linking directly to the log template so students could access it before leaving the room.

I asked them to have an explicit conversation with their advisor about AI policies and suggested they ask directly about what assistance is appropriate, how it should be documented, and what their advisor would want to know.

I also encouraged them to continue talking with their peers about their practices, because the ethical questions doctoral students face around AI are best worked out in a community rather than in isolation.

Optional: Final Poll (1 minute)

I added a closing poll asking students which persona they were most likely to try responsibly in the next month, with the option to select "none" and indicate they needed more time to think it over. The results gave a natural way to close with energy, reflecting to students what they were taking away and signaling that trying one persona thoughtfully was a reasonable next step.

Closing Message

I ended with the same message: The goal is responsible use, not perfect use (see Figure 8). Students will make mistakes and encounter gray areas that these personas do not fully resolve. The framework is a beginning, not an ending, and a foundation for ongoing ethical practice throughout their doctoral journey and beyond.

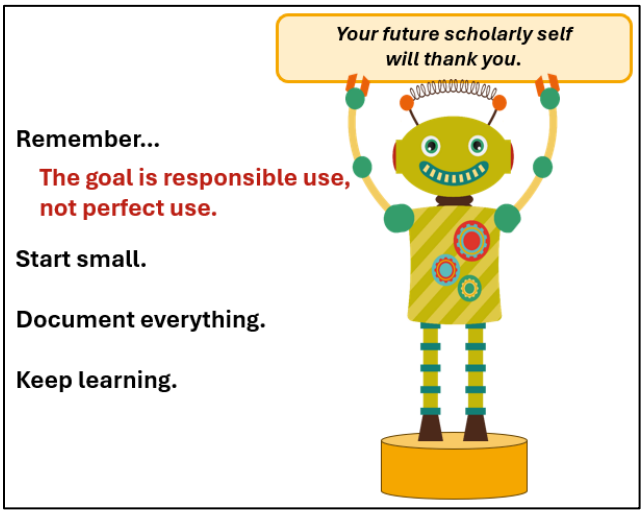

Figure 8. The goal is responsible use, not perfect use.

I stayed after the session to continue conversations and gave students my contact information for questions that came up later.

Extended Version (15-30 minutes)

If there is additional time, instructors can use the six Scenario Cards for deeper practice with the personas framework. Each card presents a realistic situation involving one or more of the five personas and gives students a chance to work through ethical decisions using the red flag questions as a guide.

In Activity 2 of the core content session, the instructor uses a single scenario to model the application of the framework. In the extended version, small groups could each receive a different scenario card and work through the analysis together before sharing with the full group. This format would add approximately 15-30 minutes to the total

session, depending on the number of scenarios used and the depth of discussion. The small group structure would also address one of the limitations of the core session: Students would benefit from debating interpretations with peers rather than simply observing the instructor model the process. This extended format has not yet been implemented but was developed in response to students consistently requesting more time for scenario practice.

Total Time: 30 minutes

- Phase 1: 3-4 minutes
- Phase 2: 1-2 minutes
- Phase 3: 15-18 minutes
- Phase 4: 8-10 minutes
- Phase 5: 2-3 minutes

Flexibility

If running behind, the instructor could abbreviate Academic Concierge (Phase 3) and shorten scenario discussion (Phase 4). Instructors should preserve the final 5 minutes for checklist distribution and closing.

Assessment

Pre-Assessment (Formative)

The opening poll served as the primary pre-assessment, revealing students' baseline understanding of AI ethics. I analyzed response patterns in real time, focusing on which types of concerns came up most often. For example, from one implementation, surface-level concerns included "The thought of using AI gives me anxiety for fear of getting kicked out of the program" and "Having the dissertation and degree taken away many years after graduating." Practical concerns included "I don't know how much is too much" and "I am concerned about my work being marked as completed by AI even when I use AI just for editing my writing." Metacognitive concerns included "At what point does the editing and revisions stop becoming my own work/thoughts?" and "Allowing it to sway my own thinking." Those responses shaped how I framed the personas discussion in that session.

Paying attention to the mix of concern types early can help instructors shape their emphasis during the personas discussion. A group where surface-level concerns come up most often may benefit from more foundational framework-building, while one where metacognitive concerns are more prevalent can move more quickly into deeper engagement during the scenario analysis.

During-Lesson Assessment (Formative)

During the session, I used hand-raising questions to reveal personal experience levels and create low-stakes opportunities for students to acknowledge their AI use without defensiveness.

The scenario poll responses were the most informative mid-session indicator and revealed whether students could apply the personas framework to a realistic situation rather than just recall its components. Responses that identified a specific persona, articulated a blind spot, and proposed a documentation strategy showed genuine application of the framework rather than just surface familiarity.

Post-Assessment (Summative)

The primary summative indicator was a comparison of pre-session poll responses with the questions and comments students raised at the end of the session. Evidence of learning showed up in several ways: Students shifted from vague concerns to specific language tied to particular personas, they referenced the framework vocabulary naturally in follow-up questions, and they asked practical implementation questions rather than conceptual ones. This shift from broad anxiety to specific, actionable thinking suggested that even a 30-minute session could move students toward more sophisticated ethical reasoning about their AI use.

Using this Lesson in Credit Courses

In credit-bearing courses, instructors could use a graded documentation practice assignment to support the development of ethical AI practices. Students could complete one substantive research task using AI assistance, such as a literature review,

writing improvement, or statistical analysis; document the process using the AI Assistance Log Template; and write a reflection addressing persona application, ethical blind spots, and the four red flag questions. The assignment could be assessed using seven criteria:

1. Log completeness,
2. Prompt specificity,
3. Verification evidence,
4. Critical use of AI output,
5. Persona application,
6. Red flag awareness,
7. Reflection depth.

This assignment could be particularly effective when students submit entries at multiple points during a semester, allowing instructors to support students as they develop their documentation practices.

Critical Reflection

The primary learning objective was consistently met in my implementation with approximately 75 doctoral students at the University of Houston's College of Education ADVANCE professional development conference. Pre- and post-poll comparisons showed students shifting from vague concerns like "Is it cheating?" to sharing specific blind spots such as "voice authenticity when using the Writing Coach."

The personas framework proved more durable than I expected. Students continued using the vocabulary weeks later without formal reinforcement. I believe this was partly due to the robots' fun and quirky representations, which made each persona appealing and memorable. Interactive polling created immediate engagement and a sense of community, and anonymity encouraged honesty about AI use that students might otherwise have been reluctant to share. The Daily AI Ethics Checklist proved useful: simple, portable, and immediately usable. Two faculty colleagues subsequently adapted the materials for courses they taught, both using the provided slides, which include a script in the presenter notes.

Psychological safety mattered more than I anticipated. The opening poll revealed the full range of student perspectives. Some responses reflected genuine fear: “The thought of using AI gives me anxiety for fear of getting kicked out of the program,” and “Having the dissertation and degree taken away many years after graduating.” Others reflected practical uncertainty: “I don't know how much is too much,” and “At what point does the editing and revisions stop becoming my own work/thoughts?” A few expressed no concern at all: “Absolutely no concerns. I am beyond excited to use AI.” The breadth of that range in a single room made it clear that a one-size-fits-all approach would not work. I found that opening with explicit reassurance that the session was about going forward thoughtfully, rather than catching anyone, was essential for engagement across that range. The non-judgmental tone throughout created the conditions for that honesty, and I believe it was what allowed students to participate openly rather than defensively.

I could also have made stronger use of the opening poll by writing the top concerns on a whiteboard and explicitly addressing each one as I moved through the personas. This strategy would have strengthened the connections between the pre-assessment and the core content. The 30-minute constraint created tension between depth and breadth that I couldn’t fully resolve. Five personas in 15 minutes meant three minutes each, which was enough to introduce the framework but not enough for students to practice applying it. One scenario was insufficient; students needed multiple practice opportunities, ideally in small groups where they could debate interpretations together.

The absence of systematic follow-up was a missed opportunity. While anecdotal evidence suggests lasting impact, I cannot confidently assess whether students changed their documentation practices or had the advisor conversations that I encouraged. The lesson also addressed individual ethics but did not sufficiently address the systemic challenges students face: conflicting advisor expectations, absent institutional policies, and evolving norms.

There are a few additions I would make in future implementations. An extended scenario library covering qualitative analysis, survey design, theoretical frameworks, grant writing, and IRB applications would give students more practice opportunities. Discipline-specific slide decks for STEM, humanities, and social sciences could be more relevant in other contexts. A brief guide for advisors, summarizing the framework and suggesting conversation prompts, would support the student-advisor conversations I encouraged. A brief follow-up at two weeks and at one month could help

assess whether students changed their practices in ways that a single session cannot capture.

Conclusion

This lesson successfully introduced graduate students to ethical frameworks for AI use, meeting its primary objective of building awareness. The personas framework proved intuitive and memorable. The lesson functions best as a catalyst for ethical thinking rather than a comprehensive solution, and it is particularly effective when followed by opportunities for practice and feedback. As AI capabilities evolve, specific examples and scenarios will need updating, but the red flag questions, the personas framework, and the emphasis on documentation remain relevant because they focus on scholarly development rather than on any particular tool or policy.

That relevance did not happen by accident. The five design decisions were central to translating situated learning and metacognitive reflection into a structure students could remember and use on their own.

1. Personas Framework
2. Interactive Polling
3. Judgment Over Rules
4. Documentation Emphasis
5. Practical Takeaways

Grounding the lesson in situated learning theory meant that ethical principles were not presented in the abstract but through recognizable use cases that students encountered in their own work. The metacognitive reflection strategies built into each phase gave students tools for ongoing self-assessment rather than a one-time set of rules. Together, these decisions translated theory into a structure that students could understand, remember, and use independently long after the session ended.

AI is not inherently good or bad, ethical or unethical. It is a powerful, flawed, and evolving tool. Like any tool, its ethical status depends on how we use it and toward what ends. This lesson equips students to make those determinations thoughtfully and in ways that support their development as scholars. The goal is not perfection. Students will make mistakes, encounter gray areas that these frameworks do not fully resolve, and face situations that require judgment beyond the scope of any lesson. That is expected and appropriate.

This lesson aims to develop students who recognize the ethical dimensions of AI use, think carefully before acting, document their practices transparently, prioritize learning over efficiency, and contribute authentically to their fields. The goal of this lesson is to support students in becoming intentional, thoughtful practitioners and scholars who can ask the right questions about their AI use, document their decisions, and develop the habits of mind that will serve them throughout their careers.

Acknowledgements

The author used Claude (Anthropic) during the initial development of this lesson to brainstorm innovative ways to present AI ethics in a format that felt approachable and engaging rather than rule-based or punitive. Claude suggested several ideas that the author evaluated, developed, and shaped into the personas framework, lesson structure, and assessment that appear in this manuscript.

Claude was also used during manuscript revision to improve readability, restructure content, and refine language in response to reviewer feedback. Claude proposed revisions that the author evaluated, refined, and incorporated at her discretion.

All instructional design decisions, implementation experience, and reflections are the author's own, as are the presentation slides and all visual and design elements in the supplemental materials, none of which were AI-generated. AI assistance was limited to brainstorming and editing of author-originated content, consistent with the JTILT AI use policy.

References

Everson, K. (2025). *Enhancing dissertations ethically with AI: A comprehensive guide*. Myers Education Press. Link to Myers Education Press

Helga-Helga. (2026). Retro hipster cute robots collection. [Illustration]. Envato. https://app.envato.com/search/graphics/79a31de9-4fe8-40fa-a8de-813ff9d3fe35

Lesh, J., & Lancaster, J. (2024). AI-Assisted academia: Unveiling doctoral students' perspectives on Dissertation in Practice innovation. *Bulletin of Educational Studies, 3*(2), 113-126. https://doi.org/10.61326/bes.v3i2.288

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology, 74*(5), 570-581. https://doi.org/10.1002/asi.24750

Parker, J.L., Richard, V.M., Acabá, A., Escoffier, S., Flaherty, S., Jablonka, S., & Becker, K. (2025). Negotiating meaning with machines: AI's role in doctoral writing pedagogy. *International Journal of Artificial Intelligence in Education,* 35, 1218–1238. https://doi.org/10.1007/s40593-024-00425-x

About the Author

Sara McNeil, Ed.D., is an Associate Professor in the Learning, Design, and Technology program at the University of Houston College of Education. Her background spans instructional design, educational graphics, and online learning, work that has shaped her current research and teaching on the responsible and ethical integration of generative AI in higher education. She focuses particular attention on how graduate students develop critical judgment about AI tools in research and academic writing. One of her current studies examines AI literacy not as a standalone subject but as an integrative framework, positioning future educational leaders to ask, across every tool and every decision:

"What can AI contribute here, and what is yours to decide?"

Sharing & Modification Permissions