

www.onehe.org

Develop your teaching your way

Guardrails, Not Gotchas: AI Ethics, Practice, and Policy That Support Real Learning

Christian Moriarty, JD MA

Professor of Ethics and Law, St. Petersburg College
Director and Treasurer, International Center for Academic Integrity

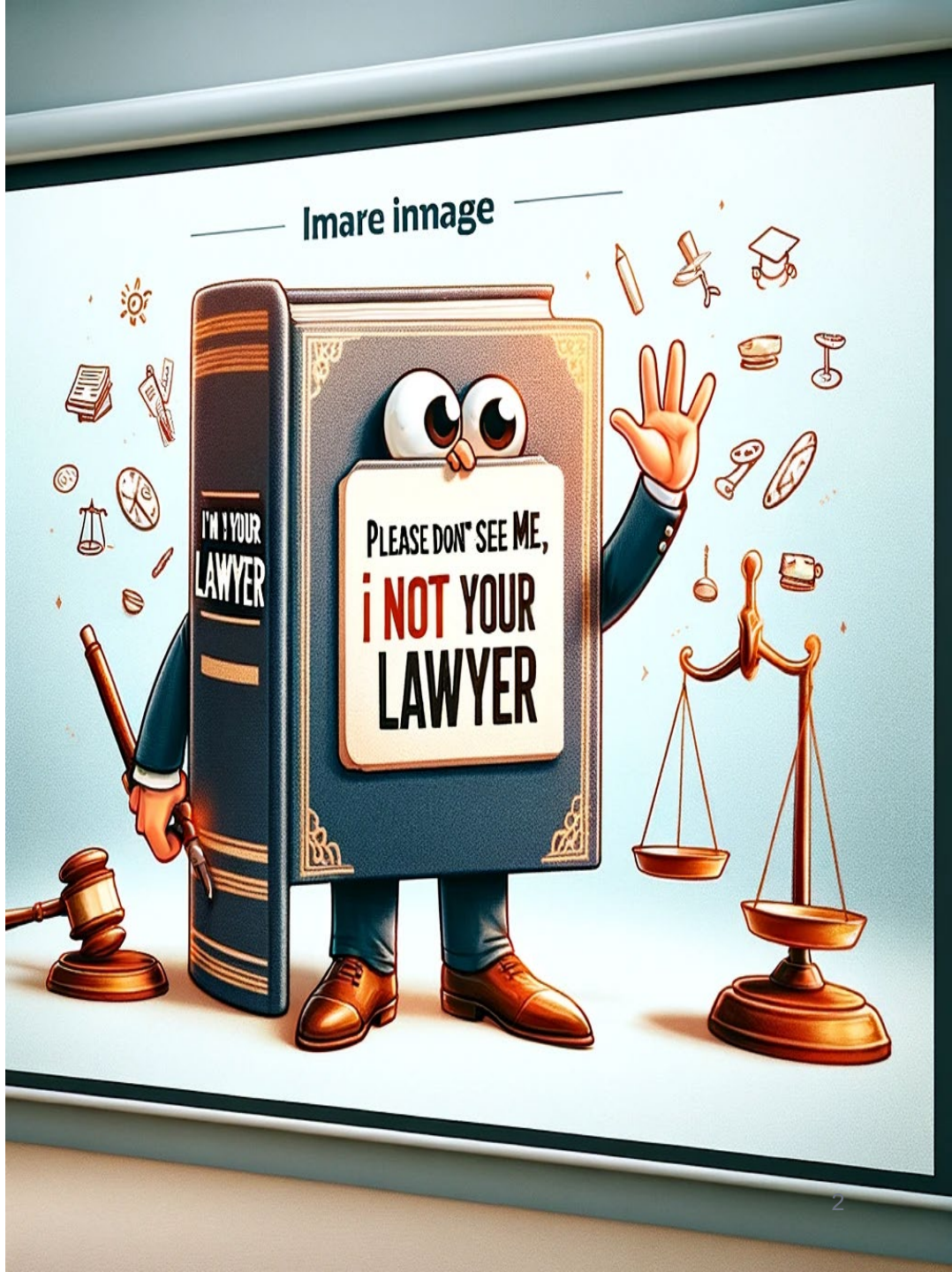
(Some pictures and slides were produced using Gemini or ChatGPT. All concepts and ideas not otherwise cited are the responsibility of the author, and AI was not used for their production.)

Warning: not your lawyer

Not legal advice. (Unless you want it, let's talk.)

See your legal department.

Don't sue me (please?)



Nearly every student uses AI, while most US campuses are still writing the rules

UK undergraduates, December 2025



In the US

32% of institutions have a campus-wide AI policy in place.

23% of faculty consider their AI policies effective.

Faculty at institutions whose policies embrace AI are twice as likely to call them effective as faculty at institutions with bans, a correlation worth weighing for anyone drafting one.

Sources: Stephenson & Armstrong (2026), Student generative AI survey 2026, Higher Education Policy Institute, n = 1,054 full-time UK undergraduates. Tyton Partners (2026), Time for class 2026: The AI tipping point, survey of US college and university faculty, administrators, and students.

Agents are moving from helping to doing the work

Agents are already in use. Agents that can log in, move through a course, and complete and submit work are already in use.

Delegation still breaks a rule (if you have it!). Unless an assignment permits it, having an agent complete the work violates integrity expectations, although it may fall under a different provision than a typical AI-use rule.

Most policies have not caught up. Policies written with AI-generated text in mind may say little about handing the work itself to an agent, including an agent logging in as the student.

Because assisting with the work and delegating it are different acts, a clear policy should address each one directly.

Detection is futile, beside the point, and unjust

Futile. Simple editing techniques measurably reduce detector accuracy, and agents make it harder still to keep pace, so detection is arguably a race institutions are unlikely to win.

Beside the point. Even an accurate catch says little about whether the student learned, and a credential ultimately has to vouch for the learning process, not just the finished product.

Unjust. Detection tends to reward fluent evaders while falling hardest on students with less AI fluency, and students report anxiety about false accusations.

Guardrails, not gotchas. If your institution uses detectors, treat them as advisory only and let process evidence govern.

Sources: Detector accuracy, Perkins et al. (2024), International Journal of Educational Technology in Higher Education. Student anxiety, Stephenson & Armstrong (2026), HEPI.

Newby v. Adelphi shows what happens when a detector score stands in for a fair process

The accusation. A professor gave a first-year student a zero after a detector scored his essay as 100 percent AI-generated. The student denied using AI and submitted results from two other detectors indicating the essay was human-written.

The process. The university's integrity officer found a violation, and the same officer then decided the student's appeal.

The ruling. The court annulled the finding, rescinded the sanction, and ordered the record expunged, largely because the university did not follow its own code (including a promised advisor and a meaningful opportunity to be heard) or meaningfully address the student's contrary evidence.

Matter of Newby v. Adelphi University, 2026 NY Slip Op 26021 (N.Y. Sup. Ct., Nassau County, Jan. 28, 2026).

Five principles hold across disciplines

Transparency. Can a student tell, before starting, what is allowed?

Fairness. Would this rule and its evidence hold up if challenged?

Accountability. Who is responsible for this output, and can they explain it?

Privacy. What data enters the tool, and where does it go?

Inclusivity. Who does this rule advantage or disadvantage?

Faculty conversations tend to go better when colleagues agree on these principles before debating specific rules.

Institutions set the floor, and faculty set the zone for each assignment

Red

Not permitted

Used when the assignment assesses the very skill AI would perform

Yellow

Permitted with conditions

Only the uses the instructor permits, and each use is disclosed

Green

Encouraged, with conditions

AI is part of the learning, and its use is still disclosed and verified

The institutional floor Shared principles, a disclosure expectation, data rules, and a fair process for resolving disputes

A student may work in the Red zone on Tuesday and the Green zone on Thursday, and when each choice is tied to an objective, that variation reflects deliberate design.

Let the learning objective choose the zone, and remember that Red is a legitimate choice

Three questions before you pick a zone

1. What must the student be able to do on their own when this assignment is done?
2. Where would AI support that learning, and where would it replace it?
3. How will I see the process, not just the finished product?

Choose Red when

- The skill itself is the objective, such as foundational writing, math, or coding fluency
- Students need a baseline of their own before they can judge what AI hands them
- The assessment has to certify what an individual can do on their own

Make it work: explain the why, build in work you can see (in class, supervised or proctored, or oral), teach the skill, and accept that a Red rule outside class is an expectation, not a guarantee.

Choosing no AI is as intentional as choosing full integration, as long as the reason is the learning objective rather than fear of the tool.

See: Stephenson & Armstrong (2026), HEPI, recommending AI-free assessments alongside assessments that expect students to use AI.

Scenario: an AI-drafted care plan for a simulated patient

A nursing student uses AI to draft a care plan for a simulated patient case and submits it with only light edits, and the finished plan reads well.

Consider

- Would your answer change if real patient data were involved?
- Where must clinical reasoning be visible in the student's own words?
- Would AI paired with instructor feedback change your answer?

Vote in the chat: **Red**, **Yellow**, or **Green** for this assignment?

Clear AI policy language covers five things

When students cannot tell what is allowed before they begin, the problem arguably lies more with the policy than with the students.

Want sample language for each zone? Reach out through ChristianMoriarty.com and I'm happy to provide it.

What. The zone for each assignment, with concrete examples of permitted and prohibited uses.

Why. A sentence connecting the rule to the learning objective it protects.

How to disclose. Which tool the student used, for what purpose, and what they kept or changed.

Data limits. Which kinds of information should never go into an AI tool.

When unsure. A prompt to ask first, and how grading decisions differ from misconduct findings.

Assessments can make student thinking visible

Version history
Drafts that show how the work developed over time

In-class checkpoints
Short tasks completed during class time

Verification tasks
Finding and correcting errors in AI-generated output

Proctored assessments
Supervised exams or tasks, in person or online, where certifying individual ability matters most

Short AI-use reflections
Brief notes on what the student asked, what they kept, and why

Oral walk-throughs
A few minutes for students to explain and defend their choices, live or by recorded video in online courses

Staged submissions
A proposal, draft, and final version, each with feedback

In the age of AI, credentialing requires assessing process rather than product.
Then again, is process ALSO product now?

Process-based design is right, and not yet enough

What we know helps

- Authentic assessment can reduce AI shortcuts
- Designing for process builds transferable skills
- Conversation over surveillance can create real accountability
- Some uses of AI genuinely support learning

What we still have to reckon with

- Students from more advantaged households are more likely to use some AI tools
- AI can now imitate an authentic voice, and even process logs
- The literacy divide shifts who succeeds, not whether the work is genuine
- All of this requires more time, energy, and money

A solution AI literacy as core curriculum, not an optional elective or an assumption, designed to reach first-generation and under-resourced students before the gap widens further.

Household advantage in AI tool use: Stephenson & Armstrong (2026), HEPI, n = 1,054 full-time UK undergraduates.

There may not be a solution, only better responses

Not all of this will work. Students will still cheat.

Why? Because it is easy to do and hard to prove.

This framework is a start, but it is not enough on its own, and pretending otherwise would be its own kind of integrity failure.

No singular solution. There is no perfect or immediate fix to the integrity and equity challenges AI raises.

A moving target. As tools improve, reliable automated detection is arguably becoming less feasible.

Sustained commitment. Progress depends on long-term investment in AI literacy, staff development, and infrastructure, not just reactive policies.

Continuous iteration. Review policies and assignments regularly as the technology changes.

Restoration, not only punishment. When fault occurs, acknowledge it, help the student learn from it, and restore them to the community.

Christian Moriarty, JD MA

ChristianMoriarty.com

Professor of Ethics and Law, St. Petersburg College

Director and Treasurer, International Center for Academic Integrity

Questions, sample policy language, or a session for your campus? Reach out through the website.

Scan for ChristianMoriarty.com



www.onehe.org

oneHE

Develop your teaching your way

Thank you!

support@onehe.org