

MODULE 5

Transparent Policies, Academic Integrity, and the Equity/Ethics

Lecture 5.1

Why Detection Fails as a Teaching Strategy

Shifting from detection to design

Four detection claims



AI-content detector

probabilistic signal



Plagiarism detector

identifies text overlap



Authorship evidence

contextual clue



Process evidence

course-designed learning
evidence

*Process evidence is the most actionable source instructors can design.
No single source should establish misconduct by itself.*

The reliability problem

Model and detector updates change performance

Editing and paraphrasing change detectable patterns

Performance varies by language, genre, and text length

A detector score is a probabilistic signal, not proof.

The inequity problem



Research has documented false positives affecting non-native English writers.



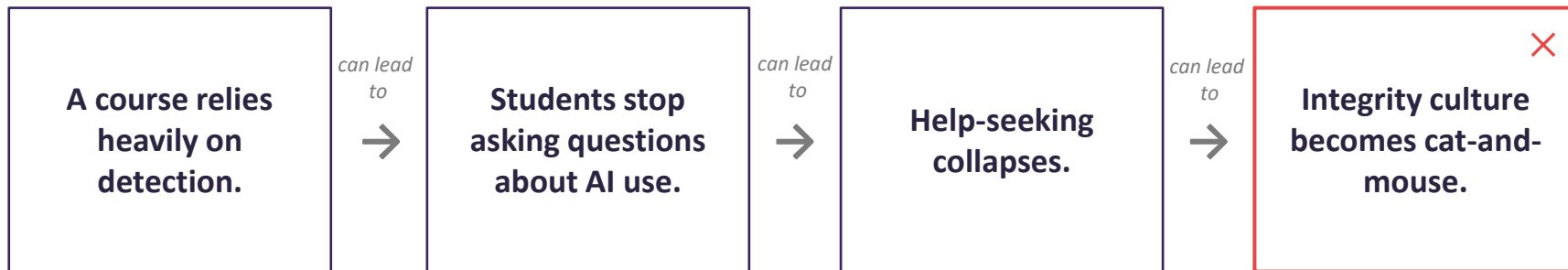
Policies must also protect students whose legitimate writing styles may be misclassified.



Policies must also protect students whose assistive workflows may be misclassified.

False positives place the burden on students to prove authorship after an accusation.

The trust problem



These risks may remain invisible until student questions decline and integrity disputes increase.

The replacement evidence stack

Process evidence built into the assignment (Module 4).

In-course conversation about AI use (low-stakes, course-internal).

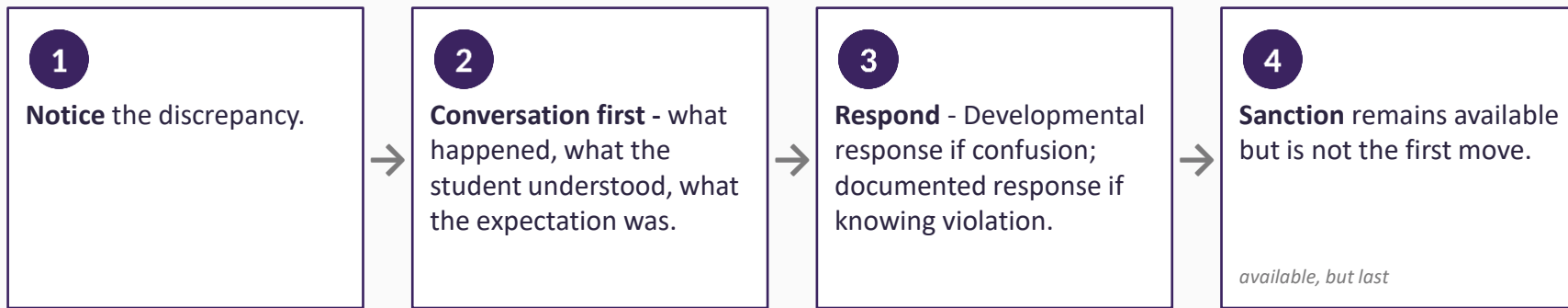
Comparison to demonstrated capability (in-class quiz, oral defense).

Student documentation required by the assignment.

Defensible evidence

*Stacked, these provide stronger evidence than any detection tool. Designed, not detected.
These sources provide context; no single layer proves misconduct.*

Restorative integrity (Eaton)



This is not "softer." It is more accurate. Conversations produce information; accusations produce defenses.

All responses must follow institutional policy, documented procedures, and student due-process protections.

Policy language: before / after

✗ Deterrence

Any AI-generated work is plagiarism and will result in failure of the assignment and reporting to the Dean of Students.

✓ Invitation

Submitting AI-produced work as evidence of cognitive work an assignment requires you to perform independently is academic dishonesty. Each assignment identifies permitted AI use, required documentation, and the work you must own.

Replace deterrence-by-threat with definition-and-invitation.

What this lecture is not saying

We are not saying integrity does not matter.

We are not saying violations should be ignored.

What we are saying

Detection scores are not the evidence integrity processes need. Design and conversation produce stronger evidence.



Elaboration micro-task *In one sentence, name the replacement evidence you would use instead of a detection score for your assignment.*