instats

B4 — Briefing for Research-Integrity Committees and Graduate Studies Committees

Responsible AI use in doctoral research is a research integrity question. Looking at AI purely through the lens of plagiarism misses central questions of validity, reproducibility, citation verification, and disclosure.

Browse the full online instrument catalogue to read or download every resource in HTML, Word, PDF, and spreadsheet formats.

About this instrument

Purpose	Give the committee that writes the institutional AI clause and hears cases a defensible framework, five decisions it owns, and five questions to ask this month.
Who it is for	Members of [COMMITTEE], such as the research integrity committee, graduate studies committee, or whatever it is called locally (including graduate research committee, doctoral board, board of graduate studies, or a joint body). Assumes no prior reading of *Responsible AI in Academic Research: A Competency Framework for Research Training*.
Time required	8 minutes to read, and 45 minutes to work through the five decisions in a meeting.
Related report sections	§1.1, §1.2, §1.3, §2.2, §2.3, §2.4, §2.5, §3.3, §3.4, §4.2, §5.3, §5.4, Appendix C, Appendix G.
Grid cells touched	D1-policy, D1-process, D2-policy, D2-process, D3-policy, D3-systems, D4-people, D4-process, D5-policy, D5-process.
Companions	**P2** disclosure standard · **G1** supervisor-student AI-use agreement · **A1** maturity self-assessment.

Table 1. Summary of the purpose, scope, and key components of this policy guidance instrument.

The argument you are most likely to disagree with

Your committee already has good machinery for one question: did this person present someone else's work as their own? Generative AI does not change or weaken that machinery. In fact, every major academic publisher concluded within about three months of ChatGPT's release that an AI system cannot be an author, because an AI tool cannot bear responsibility for the work. But four other questions now decide whether the research is actually sound. Are the citations real? Is the analysis valid? Can another researcher reproduce it? Was the tool's role disclosed specifically enough for an advisor, supervisor, or examiner (a dissertation committee member, in US usage) to assess? None of these questions is answered simply by asking who typed the words.

In the most-cited primary study on this topic, 636 generated citations across 42 literature reviews were checked against Google Scholar, Scopus, and Web of Science: 55 percent of the GPT-3.5 citations and 18 percent of the GPT-4 citations did not exist. Among the citations that did exist, 43 percent and 24 percent respectively carried substantive errors (Walters and Wilder, 2023). A cross-model study in the medical literature found fabrication rates between 28.6 and 91.4 percent (Chelli et al., 2024). A fabricated citation is not plagiarism. Nothing was copied, and no original author was wronged. It is outright fabrication, which already sits at the top of every research integrity code.

Meanwhile, the conduct that the plagiarism frame does catch, such as a graduate student polishing their own paragraph without saying so, is a disclosure omission at worst.

The broader university sector has this same scoping problem. Across 38 top-tier doctoral universities in 15 countries and jurisdictions, 15 institutions stop at what I call the plagiarism ceiling, meaning their policies treat AI only as a question of student copying. Specifically, four handle AI only inside student conduct rules, while eleven have a dedicated AI policy whose scope ends at plagiarism. Seventeen universities extend their policies into research integrity and reproducibility. Six go even further, addressing AI literacy and examiner expectations. The Australian Research Council and the National Health and Medical Research Council state the operative standard clearly: *"AI-generated content must be verified and should not replace expert opinion or judgement"* (joint policy, April 2026). Verification is a fundamental research integrity duty, not simply a student conduct rule.

What is genuinely an integrity matter, and what is not

Bring to [COMMITTEE] as an integrity matter:

1. **Fabricated content presented as real**: citations, quotations, data, or figures that do not exist or were generated rather than observed. This is straightforward fabrication under your existing integrity code, so no new category is needed. Nine major publishers place category prohibitions on AI-generated or AI-altered images in submitted manuscripts, outside a narrow exception where the AI itself is the object of study.
2. **Undisclosed use where a written standard required disclosure and the researcher was on notice of it.** That final requirement is essential: if the institution had no published standard in place and the researcher was not on notice of it, there is no formal breach.
3. **Confidential material placed in a third-party tool**: unpublished data, participant-identifiable data, a manuscript under review, or a grant application under assessment. Sixteen of eighteen surveyed publishers treat a reviewer uploading a manuscript to a public AI tool as a confidentiality breach. In addition, thirteen of the fourteen national funders surveyed restrict or prohibit assessors from doing the same with proposals.
4. **Concealment or falsification during an investigation.** This is standard misconduct, and the presence of AI changes nothing about how it should be handled.

Handle through supervision, training, and curriculum, not a hearing:

- Scoping the literature with an AI tool without verifying the sources, before the requirement to verify was explicitly written down and taught.
- Submitting a disclosure that names a tool but omits its version, the specific task performed, or what the human researcher verified, against an institutional standard that never specified those fields.
- Treating an AI system's agreement with the researcher's own argument as validation. This is known as the sycophancy problem, where an AI tool simply tells the researcher what they want to hear, and it represents a methods-training gap rather than bad intent.
- Failing to record the model version, parameters, and prompt used for an AI-assisted analysis. This is a reproducibility gap of the kind that study pre-registration was invented to close.
- Using a consumer-tier tool because [INSTITUTION NAME] provided no approved institutional alternative. That is an institutional procurement and training failure and may be mitigating context. It never excuses fabrication, concealment, confidentiality or privacy breaches, or other misconduct. The Samsung incident of March 2023, which involved three confidential leaks in twenty days followed by a corporate ban on consumer AI, clearly shows that tool choice is also an institutional responsibility.

The five decisions [COMMITTEE] owns

1. Which instrument AI cases are adjudicated under: student conduct rules, the research integrity code, or both with a clearly stated boundary. Until this is settled at your institution, every case will be argued twice.

> ☐ Yes ☐ Partial ☐ No (*Evidence: the clause and its number, the instrument it sits in, and its last review date.*)

2. The case taxonomy and its evidentiary standards. Four distinct types of cases should not share one process: undisclosed use, disclosed-but-unverified use, AI-fabricated content, and AI-assisted misconduct. Without this distinction in place, a serious fabrication case risks being treated as a routine formatting lapse.

> ☐ Yes ☐ Partial ☐ No (*Evidence: the adjudication framework, showing the standard of proof and remedy range for each case type.*)

3. The disclosure standard, and the date it starts. An institutional standard binds researchers from the date it is published and taught. It should not retrospectively reclassify conduct that was permitted when undertaken solely because the standard later changed, subject to continuing legal, ethical, contractual, confidentiality, funder, and publisher obligations. See **P2**.

> ☐ Yes ☐ Partial ☐ No (*Evidence: the published standard, the thesis or dissertation declaration form, the effective date, and the record showing that graduate students, advisors, and supervisors were notified.*)

4. What examiners may and may not do: whether an examiner may run a thesis through a detection tool, whether any part of an unexamined thesis may enter an external AI tool at all, and what the oral examination (the viva, or thesis defense) may ask about AI use.

> ☐ Yes ☐ Partial ☐ No (*Evidence: the examiner briefing note or appointment letter carrying the rule, and the relevant examination regulations clause.*)

5. What record will exist to adjudicate on. Without a clear record, every case comes down to one person's word against another. The practical options are session logs kept under a stated retention window, the thesis declaration, the supervisor-student agreement (**G1**), and pre-registered prompts and analysis plans.

> ☐ Yes ☐ Partial ☐ No (*Evidence: the retention schedule naming AI session logs, and a supervisor-student agreement drawn at random from current enrollments.*)

The five questions to ask

1. **Under which instrument would we adjudicate an AI case brought today?** *Evidence: name the policy and clause, and state its last review date.*
2. **Can our administration show us twelve months of AI-related cases classified by the four case types?** *Evidence: the case register with a case-type field. If that field does not exist, that gives you the answer.*
3. **What exactly does our disclosure standard require of a doctoral student?** *Evidence: the disclosure template and thesis declaration form side by side, showing the required fields.*
4. **What are examiners told, in writing, about AI tools and detection software?** *Evidence: the briefing note actually sent for the most recent examination.*

5. **If a researcher or graduate student disputes an allegation, what objective record do we hold?** *Evidence: the session-log retention schedule, a sample supervisor-student agreement, and the pre-registration record where the analysis was pre-registered.*

How disclosure standards work in practice

A disclosure standard is enforceable only if it clearly specifies four things: which uses require disclosure, where the disclosure appears, which specific fields it requires (the tool, the version or date, the task performed, and what the human verified), and who signs it. Established publisher practice provides useful examples for these formats: a declaration placed immediately above the references, a three-location rule affirming that the intellectual content is human, or a structured field on the submission form. Instrument **P2** in this pack provides institutional templates for these. In fact, the cleanest approach in the report's global sample is not brand new: German universities simply attached AI disclosure to the sworn declaration doctoral students already sign, allowing it to inherit that document's existing legal weight. If [INSTITUTION NAME] already uses a thesis declaration, that is exactly where this requirement belongs.

You should expect low disclosure rates at first. For instance, when the British Medical Journal introduced a structured disclosure field in April 2024, only 5.7 percent of 25,114 submissions across 49 BMJ journals disclosed AI use over the following seven months. That low rate does not by itself establish deliberate concealment. A committee that assumes otherwise will set its disciplinary response at the wrong level and lose the cooperation it actually needs from researchers and students.

A proportionate response

Your committee's response should be proportionate to what the AI use actually did to the *research*, rather than simply how much AI was used.

Finding	Default response	When it escalates
Use not disclosed, no standard in force at the time	Conversation with the advisor or supervisor, disclosure added to the record, and no formal case opened	The researcher was on notice, or denied the use when asked
Disclosure present but incomplete	Correction to the declaration, a note to the advisor or supervisor, and the item added to training materials	The omission conceals AI substitution for a human judgment task
A few unverified or wrong citations, corrected on request	Supervised verification pass over the full bibliography	The pattern spans chapters, or survives an instruction to correct
Fabricated citation, data or figure in submitted or examined work	Research integrity investigation under [INSTITUTION NAME]'s code, because no lower tier exists for fabrication	Repeated, or concealed when questioned
Confidential or identifiable material in a third-party tool	Contain first (data-protection officer, ethics approval holder, and the editor if a manuscript under review was involved), then investigate	The material was identifiable or under contractual confidentiality

Table 2. Default committee responses and escalation conditions corresponding to specific findings on AI use.

Across every row in this table, first classify the conduct and harm. Only then ask whether an approved institutional tool existed and whether the researcher was told about it. A missing approved tool is an institutional procurement and training failure and may mitigate the individual

response, but it never excuses fabrication, concealment, confidentiality or privacy breaches, or other misconduct.

Why detection tooling is not the answer

AI detection software does not work well enough to support an integrity finding. In an evaluation of seven detection tools, researchers found a baseline accuracy of only 39.5 percent, which fell to 17.4 percent under simple evasion techniques. The authors concluded that the tools *cannot currently be recommended* for determining whether academic-integrity violations occurred (Perkins et al., 2024). An earlier test of fourteen detection tools found that most performed below 80 percent accuracy, with light paraphrasing pushing the undetected rate to roughly half (Weber-Wulff et al., 2023).

These errors also fall disproportionately on those least able to absorb them. For example, a 2025 evaluation reported false-positive rates above 20 percent for writers who are not native English speakers. A false allegation against an international student or researcher carries severe visa, funding, and reputational consequences that a committee cannot easily undo.

Even if a detector were completely accurate, it would still be answering the wrong question. It only tells you whether text has the surface statistical properties of machine writing. It cannot tell you whether the citations resolve, whether the analysis is valid, whether the tool's role was properly disclosed, or whether the graduate student understands their own thesis. In other words, detection tools treat surface authorship signals as a proxy for research validity, which they are not (Appendix C, Class 11).

The instrument that answers the right question already exists, and this committee controls it: **the oral examination**. For example, the University of Toronto's School of Graduate Studies requires a doctoral student to describe and defend any use of generative AI, alongside the contents of the thesis, at the final oral examination. Similarly, King's College London instructs examiners not to upload any part of a thesis into a generative AI tool and not to use external AI-detection software when assessing it. This is a confidentiality rule as much as an accuracy rule, because an unexamined thesis is unpublished work, and uploading it is the exact act funders and publishers already prohibit for peer assessors. *If [COMMITTEE] adopts only one pair of rules from this brief, I recommend adopting that pair.*

Two dates, and one note for Australian readers

Two regulatory obligations are dated and need a named owner at your institution by [DATE].

For institutions operating in or for the EU market, where the institution is a body governed by public law or a private entity providing public services and deploys a high-risk system under Article 6(2), the EU AI Act's Article 27 Fundamental Rights Impact Assessment applies from August 2, 2026. This applies to admissions, learning-outcome evaluation, education-level assignment within institutions, and proctoring.

In Australia, the Privacy Act 1988 automated-decision-making transparency obligation commences on December 10, 2026. This requirement reaches private universities and the Australian National University, whereas most state and territory public universities are governed principally by their own jurisdiction's privacy law. Your committee might not conduct the technical assessment itself, but it does own the institutional policy language around it.

For Australian readers: the Tertiary Education Quality and Standards Agency's *Gen AI strategies for research training: Emerging practice* (June 2025) sets out what the regulator expects across induction, guidance, and training, the research process, assessment and thesis examination, and publications and grant applications. This pack complements that checklist by supplying the practical instruments to run it, along with a structured way to measure how well those instruments work.

Terminology note: "graduate student" and "doctoral student" both mean a person enrolled in a research degree, at any stage of enrollment, whatever the local name (such as PhD student or graduate student in the United States and much of Europe, research student in the United Kingdom, or higher degree by research student in Australia). "Advisors and supervisors" names one primary role across systems (advisor in North America, supervisor in the United Kingdom and Australia). "Thesis" means the thesis or dissertation. "Examiners" means the examiners, or the dissertation committee members in US usage. The "oral examination" is the final oral defense of the thesis (the viva or viva voce in the United Kingdom and Australia, or the thesis defense in the United States).

Crosswalk: brief section to report section

Section of this brief	Report section
The argument you are most likely to disagree with	The finding, §1.2, §1.3
Citation-fabrication evidence (Walters and Wilder, Chelli et al.)	§2.4, competency 1 (citation verification)
Genuinely an integrity matter, items 1–4	§5.4, §3.3, §2.3 (data residency, training-on-input)
Supervision, training and curriculum matters	§2.4, §5.3
Decisions 1 and 2 (adjudicating instrument, and the four case types)	§1.3 (Class A–D postures), §5.4
Decision 3 (the disclosure standard and its start date)	§2.2, §3.3
Decision 4 (what examiners may and may not do)	§4.2 (Dimension 1), Appendix G (examiner judgment at viva voce)
Decision 5 (what record will exist)	§2.3 (auditability via session logs), §2.5 (systems axis)
The five questions to ask	§2.5 (policy · people · systems · process), §5.4
How disclosure standards work in practice	§2.2, §3.3, §1.1 (measured disclosure rate)
A proportionate response	§5.4, Appendix G
Why detection tooling is not the answer	Appendix C, Class 11, §2.4
Two dates, and one note for Australian readers	§2.5 (regulatory pressure points), §3.4

Table 3. Mapping of specific sections in this policy brief to corresponding report sections.

instats

Michael J. Zyphur, PhD · Professor and Director, Instats · instats.org · support@instats.org

Cite the pack. Zyphur, M. J. (2026). *The Institutional AI Readiness Pack: Self-Assessment and Implementation Tools for Responsible AI in Academic Research*. Instats Policy Series. https://doi.org/10.61700/bv2nulyhht

Companion report. Zyphur, M. J. (2026). *Responsible AI in Academic Research: A Competency Framework for Research Training*. Instats Policy Series. https://doi.org/10.61700/t31oy23grr