



INSTATS POLICY SERIES · AUGUST 2026

The Institutional AI Readiness Pack

Self-Assessment and Implementation Tools for Responsible AI in Academic Research

The Institutional AI Readiness Pack and its accompanying instruments catalogue is a university-wide assessment and implementation toolkit for responsible AI in research, spanning practice, people, policy, systems, procurement, data, disclosure, and oversight. It accompanies Responsible AI in Academic Research: A Competency Framework for Research Training, which defines five dimensions of institutional readiness and the capabilities that underpin them. The pack turns that framework into instruments a university can use to get an evidence-based picture of how AI is actually used and governed across its research environment, along with ways to track that picture as it changes over time. In doing so, it connects institutional policy and strategic priorities directly with the everyday practices, capabilities, and experiences of researchers and graduate students.

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

Citation. Zyphur, M. J. (2026). *The Institutional AI Readiness Pack: Self-Assessment and Implementation Tools for Responsible AI in Academic Research*. Instats Policy Series. DOI: [10.61700/bv2nulyhht](https://doi.org/10.61700/bv2nulyhht). github.com/mzyphur/ai-readiness. ORCID: [0000-0003-3237-7892](https://orcid.org/0000-0003-3237-7892).

Companion report. Zyphur, M. J. (2026). *Responsible AI in Academic Research: A Competency Framework for Research Training*. Instats Policy Series. DOI: [10.61700/t31oy23grr](https://doi.org/10.61700/t31oy23grr). This pack operationalizes that report and refers to it throughout as *the report*.

License. © 2026 Instats. Licensed under **Creative Commons Attribution 4.0 International (CC BY 4.0)**. You may adapt, rewrite, shorten, extend, translate, and redistribute every instrument in this pack, including commercially, with attribution. You may embed the text directly into institutional policy, handbooks, AI-use agreements, and survey platforms without asking for permission. The Instats name and logo are excluded. The companion report is separately licensed CC BY-NC-ND 4.0.

Executive summary

Universities increasingly have AI policies, but many still lack an evidence-based picture of how AI is actually used and governed across their research environments. Without that picture, they cannot reliably identify gaps across policy, people, systems, and process, nor can they determine how to strengthen the institutional capabilities that enable rigorous and responsible AI use in research.

That is the gap this pack closes. The companion report, *Responsible AI in Academic Research: A Competency Framework for Research Training*, surveyed institutional AI policies at thirty-eight top-tier doctoral universities across fifteen countries and jurisdictions. It found that the typical institutional posture is an AI policy whose substantive scope ends at plagiarism. Only six of the thirty-eight published AI policies extend past research integrity into AI literacy and valid research practices. Meanwhile, the major academic publishers converged on a common position within about three months of the public release of ChatGPT-3.5, beginning with Nature and Springer Nature on 24 January 2023 and the ICMJE following in May. National research funders converged unevenly across three years, and research practice moved faster than either. The institutions that actually train researchers have not caught up.

The report's answer is a five-dimension competency framework: human-in-the-loop discipline, responsible use in practice, tooling that promotes responsible use, AI-literate humans, and an institutional benchmarking grid. The report requires policy, people,

systems, and process to operate together at the applicable maturity level. The framework applies this requirement across twenty cells, one for each combination of the five dimensions and four institutional axes. A dimension takes the level of its lowest-scoring axis, which the pack calls the binding axis. The report's conclusion on the fifth dimension is the primary reason this pack exists: no university in the sample published its own scoring against the framework, so *leading* on Dimension 5 is "a forward-looking standard that the framework's adopters can elect to populate" (the report §4.2). In this framework, the gap analysis lists improvements in order, like climbing a ladder, and each individual improvement is a rung. In other words, the report describes a rung that nobody has reached yet. This pack provides the tools for an institution to reach it.

What the pack contains. The pack contains thirteen instruments organized into four modules. The assessment suite (A1 to A5) scores the twenty cells against documentary evidence, tests those scores against what graduate students and their advisors and supervisors actually report, converts the results into concrete work packages with named owners, and reports the findings to a governing body on a single page. The briefings (B1 to B5) present the core decisions each leadership group owns in a concise ten-to-fifteen-minute read. The agreement (G1) is the flagship instrument: a thirty-minute conversation where a doctoral student and their advisor or supervisor decide, task by task, where AI may assist with the work and where judgment must remain strictly human, putting that agreement in writing. The policy library (P1, P2) supplies an adoptable institutional AI-in-research policy template alongside a disclosure standard with ready-to-use statement templates.

The [online instrument catalogue](#) lists every document and provides each available HTML, Word, PDF, and spreadsheet file.

What the assessment produces. Five dimension levels, the binding axis holding each dimension back, the overall binding dimension for the institution, a documentary data-coverage percentage, and an assessment verdict of *Final* or *Provisional*. That verdict describes the completeness of the A1 documentary assessment. It is not a claim that every survey inference has been independently validated. The assessment does not produce a single institutional score, and no adaptation should attempt to do so. The report explains that the diagnostic value of the grid lies in the dimensional pattern rather than a single summary number (§4.3). Reducing the assessment to a single number hides the specific capability bottlenecks and invites unhelpful comparisons between institutions with completely different disciplinary profiles and regulatory environments.

What it costs. Roughly twenty-two to twenty-five person-days across one quarter, plus ten minutes from each graduate student and supervisor who completes the surveys. It requires no new software, no procurement, and no external spending, unless the institution later chooses external peer benchmarking as part of pursuing the *leading* level on Dimension 5. A faster version also exists: answering four to six questions per

dimension in an hour will give you a rough sense of where you sit, while the full version produces verifiable evidence suitable for a governing council.

Why now, and not next year. Two regulatory obligations with fixed compliance dates sit inside the fifth dimension. The EU AI Act's Article 27 Fundamental Rights Impact Assessment obligation applies from 2 August 2026 to institutions within Article 2's territorial scope that are bodies governed by public law or private entities providing public services and deploy a high-risk system under Article 6(2) touching admissions, evaluation of learning outcomes, assignment of education levels, or proctoring. The Australian Privacy Act 1988 automated-decision-making transparency obligation commences on 10 December 2026. This reaches private universities and the Australian National University, whereas most state and territory public universities are governed principally by their own jurisdiction's privacy legislation. Other jurisdictions carry equivalent requirements. In both cases, the first question an auditor will ask is who owns the obligation and by what date it will be met. Answering that question requires an asset register that most institutions have not yet built.

How to get started immediately. I recommend taking two immediate steps. First, appoint a single person to own the institution's AI-readiness benchmarking cycle, publish their mandate and review schedule, and set a firm date for their first score. That appointment supplies evidence toward D5—people → nascent only if the mandate also makes that person accountable for the AI-in-research policy and its periodic review. Dimension 5 reaches *nascent* only after a complete rescore confirms that all four D5 cells qualify at that level. Second, send the five questions in briefing B1 to the five designated leaders and ask for written responses with supporting documentation attached. If those artifacts are readily available, you already hold most of the evidence base and completing the assessment becomes a straightforward scheduling exercise. If they are not available, you have an immediate answer about where the institution stands, and your gap analysis has already begun.

The framework is non-prescriptive about which level an institution should occupy. Academic scope, discipline mix, jurisdiction, and available resources legitimately differ. For example, a small humanities-focused doctoral program and a major medical research university operating across multiple jurisdictions can both defensibly sit at *established* on the same dimension with materially different implementations. Nobody outside your university is grading this exercise. There is no prize for a high score and no penalty for a low one. The only real risk is relying on an inaccurate one.

CONTENTS

Executive summary

Part 1 — Why this pack exists

- [1.1 The ladder rung nobody has reached](#)
- [1.2 The problem an institution cannot currently answer](#)
- [1.3 What this pack is, and what it is not](#)

Part 2 — The framework restated for practitioners

- [2.1 The five dimensions](#)
- [2.2 The four axes](#)
- [2.3 The four levels](#)
- [2.4 The twenty-cell grid](#)
- [2.5 The three substantive lists: modes, properties, competencies](#)

Part 3 — How to run the assessment

- [3.1 The triangulation design, and why the gap is the finding](#)
- [3.2 Who to convene](#)
- [3.3 What it costs and how long it takes](#)
- [3.4 How to keep it honest](#)
- [3.5 Six ways this goes wrong](#)

Part 4 — Reading your results

- [4.1 The cumulative scoring rule \(the maturity ratchet\)](#)
- [4.2 The dimension is the minimum of its axes](#)
- [4.3 The binding axis](#)
- [4.4 The binding dimension](#)
- [4.5 The Dimension 5 interlock](#)
- [4.6 Coverage, and the difference between Final and Provisional](#)
- [4.7 Why there is no single institutional score](#)
- [4.8 Class A–D placement is a separate measurement](#)

Part 5 — The modules

- [5.1 Module A — the assessment suite](#)
- [5.2 Module B — the briefings](#)
- [5.3 Module C — the agreement](#)

- [5.4 Module P — the policy library](#)
- [5.5 How the modules connect](#)

Part 6 — A worked example: Northgate University

- [6.1 The institution](#)
- [6.2 The twenty cells, scored](#)
- [6.3 The dimensional profile](#)
- [6.4 Triangulation: what the surveys changed](#)
- [6.5 Class A–D placement](#)
- [6.6 The three work packages Northgate should complete first](#)
- [6.7 What the profile looks like twelve weeks later](#)

Part 7 — Adapting the pack

- [7.1 The invitation](#)
- [7.2 What attribution looks like in practice](#)
- [7.3 Translation and local adaptation](#)
- [7.4 Two things to avoid](#)

Appendix A — Instrument catalogue

Appendix B — TEQSA crosswalk

- [The crosswalk](#)
- [Using the two together](#)

Appendix C — AI-assistance disclosure for this pack

Appendix D — How to cite this pack, and the sources it draws on

- [Citing the pack](#)
- [What this pack draws on](#)

Appendix E — Crosswalk to the companion framework

Part 1 — Why this pack exists

1.1 The ladder rung nobody has reached

In the companion report, §4.2 concludes its discussion of the fifth dimension with a straightforward observation:

"No university in the sample publishes its own scoring against this framework as of mid-2026, because the framework has just been published in this report, so leading on Dimension 5 is, for now, a forward-looking standard that the framework's adopters can elect to populate." (the report §4.2.)

I encourage you to read that as an open invitation rather than a qualification. The report describes what *leading* on institutional benchmarking looks like: regular self-scoring against a published framework, a designated owner for each axis of every dimension, a documented plan for applicable regulatory deadlines, published scoring with an explicit methodology, and external peer benchmarking or auditing. It then notes that no university currently does this, for the simple reason that the framework was not previously available.

That standard is available now. Establishing that position does not require a large capital investment. Publishing a scored institutional profile with a clearly stated methodology requires a committee review cycle and a web page, not an expensive program. However, this opportunity to lead will not remain open indefinitely, because academic sectors can converge quickly once expectations are defined. For example, major academic publishers established a consistent no-AI-co-author policy across the sector in roughly three months. Universities have not yet converged in the same way, which means the institutions that state a clear position now will help shape the standards that eventually emerge across higher education.

This pack provides the necessary tools to take that leadership position. Its thirteen instruments give an institution everything it needs to evaluate itself honestly, verify those scores against the experiences of researchers and students, translate the findings into concrete projects with named owners, and publish the results in a form that external peers can verify.

1.2 The problem an institution cannot currently answer

Most universities can produce an AI policy document on request. Very few, however, are likely to be able to answer the harder practical question: what actually happens when a doctoral student's AI software generates a convincing list of citations on a Tuesday afternoon?

The gap between having a written policy and understanding daily practice is the central problem this pack addresses. In my view, this problem stems from three main issues.

First, there is a fundamental category error about the nature of the problem. Treating AI use in doctoral research primarily as a plagiarism issue frames student conduct and writing as the main concern, rather than focusing on the validity, reproducibility, and ethical rigor of AI-assisted research. A fabricated citation is not an act of plagiarism because nothing was copied from another scholar and no original author was deprived of credit. It is data fabrication, which sits at the very top of research misconduct codes worldwide. In a widely cited primary study, researchers evaluated 636 AI-generated citations across 42 literature reviews against three academic databases. They found that 55 percent of the citations generated by GPT-3.5 and 18 percent of those from GPT-4 did not exist at all. Among the citations that were real, 43 percent and 24 percent respectively contained substantive factual errors. A broader cross-model study in medical literature found citation fabrication rates ranging from 28.6 to 91.4 percent. When an institution files these issues under student conduct, the problem gets routed to the wrong office, investigated under inappropriate evidentiary standards, and handled using automated detection tools that simply do not work. In testing, seven commercial detection tools showed an average baseline accuracy of only 39.5 percent, which dropped to 17.4 percent under simple evasion techniques, alongside a documented bias against non-native English writers.

Second, "responsible use" has rarely been operationalized. The phrase appears in nearly every funder mandate, publisher guideline, university policy, and competency framework in the companion report's evidence base, but its operational meaning is almost never defined in practical detail. For example, the Australian Research Council and the National Health and Medical Research Council state the requirement in a single sentence: "AI-generated content must be verified and should not replace expert opinion or judgement." That principle is entirely sound, but it does not give anyone practical instructions. It does not tell a graduate student how to handle their analysis on a day-to-day basis, nor does it tell a supervisor what specific checks to perform before approving an analysis chapter.

Third, a policy document represents only one component of institutional capability. The most common institutional response has been to publish a single AI guidance document and treat the matter as resolved. True institutional capability, however, spans four distinct axes: the policy document itself, the people designated to enact it, the technology systems that support it, and the administrative processes that integrate AI expectations into

supervision, milestones, and thesis examination. An institution can have a well-crafted policy that lacks assigned owners, retrievable records, or required workflow checkpoints. Under this framework, that configuration receives a low score, which accurately reflects its operational vulnerability.

Putting these three issues together highlights the core challenge this pack addresses: *almost any university can quote its AI policy, but very few know what their researchers and graduate students are actually doing in practice.*

1.3 What this pack is, and what it is not

It is a self-diagnostic and implementation toolkit. The instruments are structured so that every answer can be verified directly from documentary records. Each indicator explicitly names the supporting artifact required. The standard applied throughout is straightforward: *could an independent, skeptical auditor verify or refute this claim using the available documents?* If the honest answer is that an auditor would have to rely on informal assurances, the indicator is not satisfied.

It is not an external audit, accreditation, or ranking system. No outside agency receives your completed assessment. No regulator requires you to complete it, and no external body mandates its use. The framework is non-prescriptive about which maturity level an institution should target, because institutional contexts and missions vary legitimately. What the framework provides is methodological consistency, ensuring that the criteria used to evaluate maturity are transparent and applied consistently.

It is not a substitute for the companion report. This document restates the core framework in Part 2 in sufficient detail to guide working assessment sessions. However, the comprehensive evidence base (including funder policies, institutional audits, publisher standards, literature on research competencies, and detailed tool taxonomies) lives in the companion report. I recommend reading the report once before conducting the assessment, and keeping this pack on hand during working sessions.

It complements sector guidance rather than replacing it. When national regulators or sector bodies issue guidance on generative AI in research training, they generally describe *what institutions should achieve*. This pack provides the operational tools to *implement those recommendations and measure your progress*. Appendix B demonstrates this alignment in detail for the Australian higher education context, which currently offers the most comprehensive national guidance available.

Part 2 — The framework restated for practitioners

This section restates the core framework so you can conduct working sessions without needing the report open. Each subsection cross-references the corresponding section of the companion report where the detailed evidence is documented and identifies the operational method introduced by this pack.

2.1 The five dimensions

The five dimensions follow a deliberate sequence because each builds directly on the ones before it. An institution that has not established clear boundaries for human oversight cannot define what responsible use means in daily practice. Without an operational definition of responsible use, evaluating whether a software tool supports or undermines good practice is impossible. Without approved tools and operational standards, an AI literacy curriculum lacks concrete substance. Finally, without those four foundational elements, an institutional benchmarking grid has no solid evidence base to evaluate. This ordering becomes essential when resolving ties to identify an institution's binding constraint.

Key	Dimension	What it asks	the report §
D1	Human-in-the-loop discipline	Which research judgements stay human, why those, and how do you know?	§2.1
D2	Responsible use in practice	What does responsible use mean, task by task, in each of the four AI-use modes?	§2.2
D3	Tooling that promotes responsible use	Do the tools your researchers actually use make responsible use easier or harder?	§2.3
D4	AI-literate humans	Do graduate students, their advisors and supervisors, and examiners (dissertation committee members, in US usage) hold the competencies the work now requires?	§2.4
D5	Institutional benchmarking grid	Do you know where you stand, who owns each part of it, and when you look again?	§2.5

Table 1. The five evaluation dimensions, their guiding diagnostic questions, and corresponding report sections.

Dimension 1 — Human-in-the-loop discipline. This dimension reflects an institution's commitment that the core judgment steps defining academic research must remain human. Routine labor tasks may be augmented by AI, but fundamental intellectual judgments cannot be outsourced to software without compromising the scholarly integrity of the work. Labor tasks are suitable for AI assistance because errors are readily visible and straightforward to correct. Examples include polishing the prose of researcher-written text, transcribing verified audio interviews, formatting citations, or cleaning up programming code. In contrast, judgment tasks silently degrade research quality if delegated to automated systems. These include formulating research questions, interpreting unexpected results, identifying outliers, synthesizing contested literature, and determining whether findings withstand critical scrutiny. The essential institutional step is to *publish* a task-level demarcation between labor and judgment, and then enforce it through advising, supervision, and thesis examination. Appendix G of the companion report provides a starter catalogue of twenty-one research tasks that academic departments can expand for their specific fields.

The evidence anchor: a formal document classifying research tasks, an enforceable mechanism (such as an institutional research integrity code, a sworn declaration, or a thesis submission declaration form), and a documented history of regular reviews.

Dimension 2: Responsible use in practice.¹ This dimension operationalizes "responsible use" from an abstract concept into specific research workflows across four AI-use modes: *search*, *co-author*, *validator*, and *tutor*. Academic publishers have generally focused on disclosure, research funders have emphasized transparency in grant applications, and competency frameworks have focused on high-level principles. However, none of these sources specifies what a responsible literature synthesis workflow looks like, what an analysis disclosure should contain for a specific tool, or what a supervisor should verify before accepting an AI-assisted dissertation chapter. Dimension 2 provides those practical specifications.

The evidence anchor: published guidance covering each of the four modes, a disclosure template aligned with current publisher requirements, and an active supervisory or examination checkpoint that verifies disclosures against actual research practice.

Dimension 3 – Tooling that promotes responsible use. This dimension addresses an institution's commitment to procure and deploy AI tools that support responsible research practices by design. Relying solely on vendor marketing is insufficient. For instance, the companion report examines an independent evaluation of three commercial legal research tools that showed hallucination rates between 17 and 33 percent, despite vendor claims that retrieval grounding eliminated false citations. Retrieval augmentation can reduce hallucinations, but it does not eliminate them. Instead, it often shifts the failure mode from a completely fabricated citation to a misattributed real source, which is much harder to catch. Software procurement standards should therefore evaluate six observable technical properties, listed in §2.5 below.

The evidence anchor: a published procurement policy evaluating all six technical properties, a maintained inventory of AI tools used in research, clear data residency and training-on-input terms for each platform, a defined log retention schedule, and completed regulatory impact assessments.

Dimension 4 – AI-literate humans. This dimension ensures that graduate students, researchers, advisors, supervisors, and dissertation examiners develop the specific competencies required for AI-augmented research. The critical competency is not simply knowing how to operate a given software interface. Rather, it is knowing which research tasks can be safely delegated to AI, which must remain human, and how to verify any AI-generated output. Specific software tools change rapidly, but underlying research judgment remains constant. In §2.5 below, I outline six foundational competencies. These skills are essential *additions* to traditional research methods and reproducibility standards, not replacements for them.

The evidence anchor: the six competencies formally integrated into doctoral methods training, corresponding professional development for supervisors, examiner training on verification methods, and verifiable completion records for individuals rather than unmonitored optional resources.

Dimension 5: Institutional benchmarking grid. This dimension measures an institution's capacity to evaluate its standing across the first four dimensions and act systematically on the findings. Dimension 5 provides the overarching governance structure. While the first four dimensions address substantive research activities, the fifth establishes the management processes needed to review and maintain them over time. This is also where formal compliance with upcoming regulatory deadlines belongs. The operational test is whether every applicable legal obligation has an assigned owner and a scheduled compliance plan.

The evidence anchor: a formal commitment to regular benchmarking, completed assessments using a recognized framework on a set schedule, assigned owners for all twenty cells of the grid, a centralized repository of compliance artifacts, and assessment results formally presented to the university council, faculty research committees, and the research integrity board.

2.2 The four axes

Every dimension is evaluated across the same four operational axes. These are not arbitrary weightings or mere categories of evidence. Instead, they represent four practical conditions that must all be satisfied before an institutional capability can function effectively.

Axis	What it is	The characteristic failure
policy	The formal document, including its substantive scope, official approval, and scheduled review cycle.	The policy does not exist in writing, leading to inconsistent standards across different academic departments.
people	The designated roles accountable for enacting the policy, along with the training provided to them.	A formal policy exists on paper, but no one is assigned responsibility for implementing it.
systems	The supporting technology and records, including procurement controls, tool registers, and training logs.	Rules exist, but compliance records cannot be retrieved to answer internal questions or external audits.
process	The administrative workflows that integrate AI checks into supervision, milestone reviews, and examination.	Comprehensive documentation exists centrally, but it never impacts daily research or student supervision.

Table 2. The four operational axes with their definitions and characteristic institutional failure modes.

These four axes apply across all jurisdictions. While specific local regulations may vary, these structural requirements remain consistent.

2.3 The four levels

The maturity levels describe verifiable institutional practices rather than aspirational goals.

Level	Definition
Absent	The institution has not established a formal position on the dimension. The relevant policy, designated role, technical system, or administrative process does not exist.
Nascent	A position has been articulated but not fully operationalized. For example, a policy exists but lacks designated owners or processes, or owners are named but lack supporting systems.
Established	Policy, people, systems, and processes are all fully in place and operating on a regular, reviewable schedule. The institution can readily demonstrate who does what, with which tools, and through which workflows.
Leading	The institution meets all <i>established</i> requirements and provides verifiable evidence of impact, such as published evaluation metrics, external peer audits, and documented continuous improvement cycles.

Table 3. Definitions and institutional characteristics of the four assessment maturity levels.

Two key principles govern this scale. First, the scale is strictly **cumulative**. In the full maturity grid, every *leading* cell begins with the requirement "All of *established*, plus...", meaning an institution cannot be classified as *leading* if it fails to satisfy the baseline requirements for *established*. Second, the report defines the *established* level conjunctively, requiring policy, people, systems, and process to be in place simultaneously. The pack applies that rule by setting each dimension at the maturity level of its lowest-scoring axis.

2.4 The twenty-cell grid

For this pack, the core unit of assessment is the intersection of a dimension and an operational axis. Evaluating five dimensions across four axes yields twenty distinct cells, with each cell scored at one of the four maturity levels.

	policy	people	systems	process
D1 Human-in-the-loop discipline	D1-policy	D1-people	D1-systems	D1-process
D2 Responsible use in practice	D2-policy	D2-people	D2-systems	D2-process
D3 Tooling that promotes responsible use	D3-policy	D3-people	D3-systems	D3-process
D4 AI-literate humans	D4-policy	D4-people	D4-systems	D4-process
D5 Institutional benchmarking grid	D5-policy	D5-people	D5-systems	D5-process

Table 4. Matrix of twenty assessment cells formed by intersecting five dimensions and four operational axes.

These cell codes are used consistently across all instruments in the pack, allowing you to link survey items, audit indicators, policy clauses, and action items directly to the same cell. Within each cell, individual indicators follow a standard naming convention, such as `D1-policy-N1`, `D3-systems-E2`, or `D4-people-L1` (identifying the cell, the target level where N is nascent, E is established, and L is leading, and an indicator number). The *absent* level has no standalone indicators because a cell is classified as absent whenever its nascent requirements are not met.

The summary grid below outlines the criteria for each dimension across the maturity levels. The *leading* column highlights what an institution must demonstrate in addition to the *established* baseline.

Dimension	Absent	Nascent	Established	Leading (adds)
D1 Human-in-the-loop discipline	No published institutional position, leaving task boundaries to individual supervisors.	Stated as a broad principle in policy, but without a detailed task-level demarcation.	Published task demarcation, integrated into methods training, and verified in supervision and examination.	+ published evaluation scores, outcome data, and external peer benchmarking.
D2 Responsible use in practice	Limited to standard student conduct rules, with no specific guidance on AI research use.	A generic disclosure template exists, but lacks operational guidance for the four AI-use modes.	Comprehensive rules for all four modes, actively taught, disclosed, and systematically verified.	+ annual compliance data and published sector-level benchmarking.
D3 Tooling that promotes responsible use	No procurement review, leaving software choices entirely to individual researchers.	Procurement review exists but relies on vendor claims, with inconsistent citation verification checks.	Formal procurement standard covering all six properties, regular output auditing, and completed rights assessments.	+ scheduled technical audits, independent validation of vendor claims, and an ongoing improvement cycle.
D4 AI-literate humans	Treated merely as informal software skills, with no structured curriculum or training.	Isolated training options exist (such as optional library guides), but remain voluntary and uncoordinated.	The six core competencies are formal curriculum outcomes, with mandatory training for supervisors and examiners.	+ published evidence on provision, reach, and selected aligned performance outcomes, with a documented curriculum review process.

Dimension	Absent	Nascent	Established	Leading (adds)
D5 Institutional benchmarking grid	No institutional benchmarking, with unassigned operational axes and unmanaged regulatory deadlines.	Benchmarking recognized in principle, but only the policy axis has an assigned owner.	Regular self-assessment on a set schedule, named owners for all axes, a regulatory plan, and committee oversight.	+ published assessment results, external peer reviews, and contributions to sector standards.

Table 5. Maturity criteria for each dimension across absent, nascent, established, and leading levels.

Appendix A of the companion report expands every cell into its complete set of observable criteria. In this pack, Instrument A1 translates those criteria into sixty-eight concrete indicators that an institution evaluates using documentary evidence.

2.5 The three substantive lists: modes, properties, competencies

Three core lists contain the practical substance of Dimensions 2, 3, and 4. They appear throughout this toolkit and are summarized below for convenient reference during assessment sessions.

The four AI-use modes (Dimension 2).

Mode	What responsible use requires
search	Use tools for initial scoping. Verify every identified citation against the primary source before citing it. Evaluate tools based on empirically measured error rates rather than vendor marketing claims.
co-author	Allow AI assistance for refining researcher-generated text, such as stylistic polishing, formatting, or length adjustment. Prohibit AI generation of text that the researcher cannot independently explain and defend. Disclose AI assistance in accordance with target publisher standards.
validator	Use AI critiques as prompts for critical thinking, not as automated validation that replaces human peer review. Conduct validation using models from a different family than the model used to generate the original text. Recognize that an AI agreeing with your argument is often a sign of sycophancy rather than scientific validity.
tutor	Prioritize corpus-grounded tutoring tools (retrieving information from approved reading lists or institutional repositories) over unconstrained chatbots. Require graduate students to demonstrate independent mastery of core concepts before treating AI-assisted progress as evidence of learning.

Table 6. The four AI use modes and requirements for their responsible application.

The six tool properties (Dimension 3): procurement criteria.

1. **Verifiable citation and resolvability**, tested on representative sample queries during procurement rather than accepted from vendor documentation.
2. **Data residency and training-on-input controls**, requiring enterprise-managed or locally hosted environments with training on user prompts disabled for any confidential or unpublished research data.
3. **Uncertainty reporting**, providing explicit confidence metrics as a standard output rather than masking uncertainty behind confident text generation.
4. **Reproducibility at a known model version**, including date-stamped model checkpoints or weights, fixed parameters, and defined stability tolerances for non-deterministic tools.
5. **Auditability**, supported by secure retention of user prompt and session logs for a defined retention period.
6. **Open-source and local deployment options**, where required by research data sensitivity or computational reproducibility needs.

The six competencies (Dimension 4).

1. **Citation verification:** verifying every citation produced in an AI workflow against the original primary source before publication.
2. **Model-and-parameter specification:** documenting the model version, temperature, random seed (where exposed), prompt structure, and stability testing for any AI-mediated analysis.
3. **Prompt-as-fork-in-the-garden discipline:** recognizing prompt selection as a researcher degree of freedom, managed through pre-registered prompt templates and prompt-sensitivity evaluations.
4. **Model-heterogeneity in adversarial review:** ensuring that automated critique utilizes models from different model families, avoiding self-evaluation where the same underlying parameters generate both the work and the critique.
5. **Sycophancy detection and human-as-verifier discipline:** treating uncritical AI agreement as a potential failure mode, and viewing AI disagreements as flags requiring human intellectual evaluation rather than automated resolution.
6. **Structured failure-mode reporting:** documenting the evaluation protocol, threat model, testing setup, and failure-discovery rates alongside AI-assisted research results.

The unifying principle across all six competencies is rigorous validity reasoning. Researchers must treat AI systems as tools with specific, predictable failure modes (such as hallucination, prompt sensitivity, non-determinism, model drift, and sycophancy) and apply the same critical scrutiny they traditionally apply to p-hacking, publication bias, and other methodological threats.

Part 3 — How to run the assessment

3.1 The triangulation design, and why the gap is the finding

This assessment collects three independent perspectives across the twenty cells, and the primary diagnostic value lies in analyzing where these perspectives disagree.

A1: What leadership can document. Instrument A1 evaluates what the university has formally documented, approved, published, and reviewed. Its evidence base consists of concrete artifacts, such as policies, committee minutes, form templates, training records, procurement contracts, and regulatory impact assessments. It is completed by the administrative leaders who manage those records.

A2: What graduate students report. Instrument A2 asks doctoral students about their actual experiences: what guidance they received, what training was provided, what their supervisors discussed with them, what forms they completed, and what occurred during milestone reviews. Students experience the daily reality of these policies. A2 omits the five policy-axis cells because whether a policy exists is a documentary question answered by A1. Instead, A2 answers whether that policy actually reached the student body.

A3: What supervisors report. Instrument A3 gathers corresponding data from faculty: what training they completed, what standards they expect from students, what checks they perform, and what records they maintain. Twelve of its twenty-six questions address topics also covered in A2. Treat them as direct cohort comparisons only after checking that the referent, time period, denominator, and response scale align.

None of these three instruments can stand alone. A1 establishes what the institution can support from documentary records. A2 and A3 provide criterion-specific respondent evidence only for predeclared claims that are directly aligned with an item and observable to that respondent group. They are anonymous cohort surveys, not ground truth and not matched student-supervisor dyads. The framework defines the *established* level as having policy, people, systems, and process all in place **and operating on a regular schedule**. Document review alone cannot establish every recipient-facing experience, while respondent evidence cannot establish that a policy or system exists.

When these three perspectives diverge, three core reconciliation rules apply (detailed in Instrument A4):

- **Only predeclared, directly aligned respondent evidence may test a criterion at *established* or above.** The item must measure the same observable claim, population, referent, and time period. Training exposure, knowledge, confidence, and contextual items remain diagnostic unless a separate criterion explicitly calls for them.
- **Student evidence takes priority only for the same recipient-facing experience.** If supervisors report that a discussion occurred but the intended student recipients report that it did not reach them, the student evidence governs that recipient-facing criterion. For policy facts, system facts, supervisor-development records, and other differences without the same referent, report the two cohorts side by side and adjudicate the discrepancy. Neither cohort receives general priority.
- **When surveys show widespread local practice that lacks formal documentation in A1, retain the A1 documentary level and flag a practice/codification gap.** Respondent evidence does not promote a cell. Unwritten practices cannot be audited, cannot be defended to external regulators, cannot survive staff turnover, and cannot be applied equitably across departments.

Triangulation identifies five reporting statuses: a *paper gap* (where formal policy runs ahead of actual practice), a *practice gap* (where good practice exists locally but has never been codified in policy), a *transmission gap* (where supervisors report providing guidance

but students report not receiving it), an *aligned* status (where the available sightlines agree), and *documentary-only / not tested* where no directly aligned respondent item tests the criterion.

3.2 Who to convene

This assessment will fail if delegated to a single administrative office, because no single department holds all the necessary evidence. In fact, four of the twenty cells fall outside the research portfolio entirely. I recommend convening a working group consisting of the following roles or their local equivalents.

Role	What they hold
Executive Sponsor (e.g., Vice-President for Research, Deputy Vice-Chancellor Research, or Provost)	Chairs the initial session or formally delegates the chair.
Dean of Graduate Studies or Associate Dean (Doctoral Training)	Admissions, milestone reviews, supervisory allocation, and thesis examination records.
Director of Research Integrity	Institutional research policies, integrity training records, and case adjudication histories.
Associate Dean for Research (from a major faculty or school)	Practical disciplinary evidence on whether central policies reflect local reality.
Lead for Doctoral Supervisor Development	Supervisor training curricula, attendance records, and professional development resources.
Academic Integrity Officer	Misconduct reporting data and current classifications of AI-related cases.
Research Librarian / Information Specialist	AI literacy workshops, citation verification instruction, and reference management support.
Chief Information Security Officer or delegate	Enterprise software tenancies, log retention configurations, and software registers.
Data Protection Officer / Privacy Officer	Data privacy assessments, records of processing, and regulatory compliance timelines.
IT Procurement Lead	Software evaluation processes, vendor contracts, and enterprise licensing terms.
Doctoral Student Representative	Student perspectives across the entire session, ensuring transparent evaluation.
Independent Facilitator	Guides the scoring process objectively without a vested interest in the outcome.

Table 7. Recommended working group roles and the institutional evidence or responsibilities each role holds.

Drafting note on group size. A working group of eight to twelve people is manageable for a half-day session, whereas larger groups often become unproductive. If certain roles overlap at your institution, combine them. If a specific role does not exist, record that absence as a finding, as several indicators in Dimensions 3 and 5 explicitly evaluate whether these responsibilities are assigned.

3.3 What it costs and how long it takes

Step	Who	Effort	When
Appoint the benchmarking lead, publish the mandate and schedule	Executive Sponsor	1 hour	Week 1
Assemble supporting documentation against A1	Benchmarking lead and research analyst	~10 person-days	Weeks 2 to 6
Administer surveys A2 and A3	Graduate School administrator	~2 person-days (10 min per respondent)	Weeks 3 to 6 (parallel)
Score the twenty cells in a facilitated session	Working group (8 to 12 participants)	Half-day session (~5 person-days total)	Week 8
Perform gap analysis and assign action items (A4)	Core working group	Half-day session (~3 person-days total)	Week 10
Prepare governing board summary and public position (A5)	Benchmarking lead, signed by Sponsor	~2 person-days	Weeks 11 to 13

Table 8. Implementation schedule showing steps, responsible roles, estimated effort, and expected timeline.

In total, this requires roughly twenty-two to twenty-five person-days of effort over one academic quarter, alongside ten minutes from each survey respondent. It requires no new software investments or external consulting fees, unless the university decides to pursue external peer benchmarking as part of qualifying for the *leading* level on Dimension 5.

The facilitated scoring workshop takes approximately four hours: fifteen minutes to review the evidence rules, ten minutes to confirm the scope of assessment, 175 minutes to evaluate the five dimensions, a fifteen-minute break, fifteen minutes to determine the policy classification (Classes A to D), and ten minutes to assign owners and deadlines to any unresolved items.

Confirm and record the unit of assessment before beginning. By default, evaluate the central university administration unless you explicitly decide otherwise. Many institutions have pockets of excellent practice within specific departments while central governance remains weak. Averaging these together produces an uninformative score that describes neither setting accurately. If an individual school has developed local tools that the central administration lacks, record those tools in the notes column and score the central institution based on central evidence. Individual schools can also run the assessment on their own, and comparing a department's profile against the university-wide profile often provides valuable insights.

3.4 How to keep it honest

Four core rules will keep the assessment rigorous and productive.

The evidence rule. Acceptable evidence includes published policy documents with explicit clause references, live system workflows or forms, official training completion registers, formal committee minutes, signed administrative procedures, dated register entries, official reports delivered to governance bodies, contract clauses, and completed regulatory impact assessments. In contrast, informal intentions, unapproved drafts, undocumented pilots, personal recollections, one-off presentation slides, undated web pages, and vendor marketing claims do not count as evidence.

Park items when documentation is missing rather than guessing. If supporting documentation is not available during the session, mark the indicator as *parked* rather than assuming a positive answer. Only the designated holder of the artifact should answer for that item. If participants disagree about what a document states, examine the document directly. A thorough scoring session will generate detailed documentation, so expect to record explanatory notes for roughly one out of every five indicators. Parked items should be resolved within two weeks following the session. Any item that remains unresolved stays unanswered and reduces documentary coverage. It contributes no points and blocks confirmation of its rung, but it is not evidence that the underlying practice is absent.

Avoid the "Partial" trap. In my experience, *Partial* is the response most frequently misused because it feels comfortable. You should only use *Partial* when a capability genuinely exists but does not fully satisfy all specified criteria. Examples include a published policy that addresses only two of the four AI-use modes, supervisor training that is implemented

in only one faculty, or an asset register that has not been updated in eighteen months. **A 'Partial' response never advances an institution to a higher maturity level.** A maturity rung is achieved only when every indicator within it is verified as *Yes*. If a rung contains some *Partial* ratings and no *No* or unanswered items, it is classified as *in progress*. This is valuable for internal planning, but it does not raise the cell's official level. Above all, do not use *Partial* to represent future intentions. An unfulfilled intention is recorded as a "No".

Why guessing high (optimistically) defeats the purpose. Inflating your scores undermines the value of this assessment in three distinct ways. First, the primary output is an operational action plan. Every "No" and "Partial" response becomes a specific project on the improvement ladder in Instrument A4. If you artificially inflate an answer, you remove that project from your plan, meaning the underlying gap will not receive attention or funding. Second, the resulting profile is designed to diagnose structural constraints rather than serve as an institutional report card. Inflating the score of your weakest dimension conceals the exact bottleneck holding back your entire institution. Third, published maturity claims are verifiable. If an institution publishes an exaggerated rating, it creates an inaccurate public claim that cannot be substantiated upon inspection. It is far safer, and much more useful, to publish an honest *nascent* rating accompanied by a dated action plan than to claim an *established* rating without the documentary evidence to support it.

3.5 Six ways this goes wrong

1. **Delegating the entire assessment to a single person or office.** This is the most common failure mode. No single office holds the evidence across all twenty cells. An assessment completed by the research office alone will score its own areas accurately and rely on assumptions for everything else. You must convene the cross-functional group.
2. **Scoring aspirational goals rather than current reality.** Evaluating the institution you hope to build rather than the one currently operating produces an unhelpful assessment. A clear sign of this problem is a scoring session that produces no parked items and no explanatory notes.
3. **Treating the self-assessment as a high-stakes performance audit.** If participants believe the results reflect negatively on them personally, they will defend their departments rather than report candidly, leading to an artificially inflated profile. Make clear at the start that the framework is non-prescriptive, that no external body evaluates the score, and that the goal is to build a practical work plan.
4. **Skipping the student and supervisor surveys (A2 and A3).** Relying solely on A1 produces a documentary review without testing selected recipient-facing and operational claims against respondent evidence. Completing A1 alone is an acceptable initial step under tight time constraints, but the resulting profile should be described as documentary only.

5. **Averaging scores across operational axes.** Averaging an absent axis with an established one produces an artificial middle score that conceals the exact bottleneck the framework is designed to identify. A dimension's score is always determined by its lowest axis score.
6. **Procuring software before defining institutional rules.** Dimension 3 evaluates whether tools support responsible research practices as defined in Dimensions 1 and 2. Procuring AI software before establishing use policies means evaluating tools against vendor marketing rather than institutional research requirements.

Part 4 — Reading your results

4.1 The cumulative scoring rule (the maturity ratchet)

Each cell contains indicators across three sequential rungs: nascent, established, and leading. A rung is **met** when every indicator on it is verified as *Yes*. It is classified as **in progress** when at least one indicator is *Yes* or *Partial*, none is *No* or unanswered, and the full rung is not yet met. A rung is **not met** when documentary evidence supports at least one *No*. It is **unresolved** when an unanswered indicator prevents a determination and no documented *No* already establishes that the rung is not met.

An institution's level within a cell is the highest rung for which that rung **and every preceding rung** is fully met.

Cell level	Condition
Absent	The nascent rung is documented as not met or is in progress.
Nascent	The nascent rung is met, but the established rung is not met.
Established	Both the nascent and established rungs are fully met.
Leading	All three rungs (nascent, established, and leading) are fully met.

Table 9. Cumulative scoring conditions required to achieve each maturity level within an assessment cell.

This cumulative structure operates like a ratchet, ensuring that an institution cannot jump ahead without establishing the necessary foundations. In Appendix A of the companion report, every *leading* cell begins with the requirement "All of *established*, plus...". Consequently, satisfying a leading indicator while failing an established one does not raise the cell's maturity level. You should still record the leading practice in your evidence

register because it provides useful context and will shorten your implementation work later, but it does not move the institution's location within a given cell.

If an unanswered indicator prevents the nascent rung from being demonstrated, mark the cell **Unresolved/not demonstrated** alongside the provisional documentary profile rather than assigning *absent* on the basis of a blank. This is an evidence-status notation, not a fifth maturity level. For an unresolved higher rung, retain the highest lower level fully demonstrated and mark the next rung unresolved.

4.2 The dimension is the minimum of its axes

This pack assigns each dimension the lowest score among its four constituent axes. It is not the mean, not the median nor the mode, and not the highest score. This rule reflects the report's requirement that policy, people, systems, and process must all be in place and operating on a regular schedule. Therefore, a dimension with three cells at *established* and one at *nascent* remains at *nascent* for that dimension. Advancing along a dimension requires every cell to reach that level.

If any axis cell is **Unresolved/not demonstrated** at the nascent rung, report the dimension as **Unresolved/not demonstrated** in the provisional profile and list the unresolved axes. Do not assign a binding axis or binding dimension until the missing evidence is resolved, unless another documented cell already fixes the dimension at a lower level independently. This is separate from documentary coverage. A coverage verdict of Final does not make an unresolved maturity result final.

In my experience, this is the rule most frequently violated when institutions adapt maturity rubrics, and doing so undermines the diagnostic value of the assessment. Calculating an average turns an absent capability into a misleading, comfortable middle score, effectively hiding the exact institutional bottleneck the grid is designed to uncover.

4.3 The binding axis

This pack calls the lowest-scoring axis within a dimension the binding axis. It represents the specific bottleneck holding that dimension back. I recommend naming the binding axis whenever you report a dimension score. For example, stating that "Dimension 3 is at *nascent*, with a binding axis of *systems*" provides clear, actionable information, whereas simply stating that "Dimension 3 is at *nascent*" does not.

If two or more axes share the lowest score, all tied axes are classified as binding and should be reported.

Identifying the binding axis is one of the most practical outputs of the assessment because it translates an evaluation score into a concrete operational task. The four axes fail in distinct ways and require different corrective actions:

Binding axis	What it means	The remedy
policy	Staff and students are conducting research without a formal policy defining institutional expectations.	Formalize existing good practices into policy. This is typically a six-to-ten-week drafting task and yields a better policy than drafting from scratch.
people	A formal policy exists, but no designated role has been assigned to implement and oversee it.	Assign formal ownership to a specific role. This requires no capital expenditure and quickly unblocks progress across multiple cells.
systems	Policies and roles exist, but compliance data and tool inventories cannot be systematically retrieved.	Construct the necessary data registers. This is often the most time-consuming task and should be started early since other reporting depends on it.
process	Central documentation exists, but requirements are not integrated into supervision, milestones, or examination.	Embed requirements into existing milestones (such as thesis proposals or annual reviews) rather than creating separate administrative burdens.

Table 10. Operational meanings and recommended corrective remedies for each binding axis failure.

Four characteristic institutional profiles tend to recur, each suggesting a distinct strategic priority:

- **Uniformly absent or nascent across all five dimensions.** This is a common starting point for universities beginning structured AI governance. The priority is to sequence work in dimensional order, focusing initial efforts on establishing foundational policies and assigning responsible roles.
- **Policy running significantly ahead of all other axes.** This is the most common pattern across the sector: a comprehensive AI policy published on a website, but lacking assigned owners, retrievable records, or milestone checkpoints. In this situation, compliance is usually limited to individual departments that take local initiative, and misconduct cases are evaluated against rules that students were never taught. The solution is not to write more policy. The institution is already proficient at drafting documents, and creating another policy will feel like progress while changing very little in practice.

- **Strong governance in Dimensions 1 and 2, but weak technical controls in Dimension 3.** In this profile, task demarcations and use rules are well defined, but the software tools provided to researchers do not support those standards. Core principles cannot be implemented at scale until robust procurement standards and technical registers are established.
- **Extensive departmental innovation alongside weak central governance.** This decentralized pattern frequently appears in large research universities. Central governance scores poorly, but the university as a whole possesses significant pockets of good practice. The strategic task is to document these local innovations and adopt them as institution-wide standards.

4.4 The binding dimension

The binding dimension is the lowest-numbered dimension sitting at the institution's lowest maturity level. This prioritization is based on the functional sequence of the framework: the five dimensions represent a chain of dependencies. For example, if both D2 and D4 sit at the minimum level, D2 is designated as the binding dimension because developing an AI literacy curriculum (D4) requires first establishing clear rules for responsible research use (D2).

Evaluating the binding dimension first identifies your primary institutional constraint. Directing resources toward other dimensions before addressing the binding dimension will yield limited practical returns.

There is, however, an important practical distinction to keep in mind, as explained in Instrument A4. The binding dimension identifies the primary institutional constraint, but it does not always represent the *first action item to complete*. Rungs in Dimension 4 (such as curriculum approvals and supervisor development programs) often require extended academic-year governance cycles, and their success cannot be formally verified without the records systems established in Dimension 5. When the rungs of a binding dimension require long timelines, it is often sensible to implement faster foundational tasks in other dimensions (such as establishing an evidence register) while launching the longer curriculum initiatives in parallel.

4.5 The Dimension 5 interlock

The companion report identifies a fundamental consistency requirement: an institution cannot defensibly claim an *established* rating on Dimensions 1 through 4 while remaining at *nascent* on Dimension 5 (§4.1). This is because maintaining verifiable documentation and regular oversight is itself necessary to demonstrate that practices are actively operating rather than merely asserted.

This requirement is applied as a formal validation check. **If Dimension 5 is scored below *established*, any claim of an *established* or *leading* rating on Dimensions 1 through 4 must be recorded as *unassured*.** Unassured ratings may be used for internal strategic planning, but they should not be published externally, quoted to research funders, or submitted for academic accreditation until Dimension 5 reaches *established*.

This check addresses a very common finding: universities that have developed genuine, effective practices in Dimensions 1 and 2 often cannot substantiate them to an external auditor because no central office maintains the verification records.

4.6 Coverage, and the difference between Final and Provisional

Documentary data coverage represents the proportion of the sixty-eight A1 indicators that are definitively answered as *Yes*, *Partial*, or *No* based on verifiable documentation. Parked or unanswered items do not count toward data coverage. An inability to locate evidence is not proof that the underlying practice is absent.

- **Documentary coverage of 80 percent or higher:** the A1 assessment receives a verdict of **Final**, reported alongside the exact coverage percentage.
- **Documentary coverage below 80 percent:** the A1 assessment receives a verdict of **Provisional**. The institution reports its provisional levels, notes the specific cells containing unanswered items, and establishes a target date for completing the documentary review.

A Provisional rating is a normal and constructive outcome for a first-year assessment. It should be reported transparently as a baseline measurement of how readily the university can retrieve its own governance records, which provides immediate, actionable feedback for improving systems in Dimension 5.

Institutions should also establish a minimum response threshold for surveys A2 and A3 before fielding them, below which survey findings are treated as indicative rather than level-changing. One possible starting rule is a response rate of 25 percent of the eligible population, with no individual school contributing fewer than ten responses. **I strongly recommend recording these response thresholds before opening the surveys rather than adjusting them after data collection.**

Measurement and validation status. This pack is a criterion-referenced, source-anchored operational toolkit. It has not been field-calibrated as a psychometric scale. The maturity categories are ordinal, and the distance between adjacent levels should not be treated as equal. Training exposure and self-reported confidence are not proof of individual competence without a direct, aligned performance measure. Longitudinal comparisons require the same assessment unit, instrument version, survey populations, fielding

window, and coding rules. The pack should not be used to rank institutions or interpret score differences as precise quantities.

4.7 Why there is no single institutional score

This toolkit does not calculate an overall institutional score, percentage, star rating, or letter grade, and I strongly advise against adapting it to do so. In the companion report, analysis of institutional case studies showed that readiness is almost never uniform across all five dimensions. As noted in §4.3 of the report, the primary diagnostic value of the framework lies in identifying the specific pattern across dimensions rather than generating a single summary number. For instance, the worked example of a UK university scored near the top of the sample on Dimension 1 (human-in-the-loop discipline) while remaining at *nascent* on Dimension 3 (procurement and tooling). Calculating a single average score would erase both of those critical findings and describe an institution that does not exist. More importantly, an aggregated score conceals the specific operational bottlenecks that university leaders need to address. Identifying those precise bottlenecks is the entire practical purpose of this assessment.

A single institutional score is also misleading for university governance. It encourages unhelpful comparisons between institutions that have completely different academic disciplines, student demographics, and regulatory exposures. It also tempts leadership to optimize a single overall number rather than fix the specific structural weaknesses identified in the weakest cells.

The assessment produces exactly six reportable outputs:

1. Twenty recorded cell levels.
2. Five individual dimension levels.
3. The named binding axis for each dimension.
4. The overarching binding dimension for the institution.
5. The assurance status derived from the Dimension 5 interlock check.
6. The final verdict status (*Final* or *Provisional*), alongside the data coverage percentage.

If senior leadership or a governing council asks for a single summary metric, provide them with the five dimension levels and the binding dimension. These can be presented clearly on a single line.

4.8 Class A–D placement is a separate measurement

The companion report also classifies university AI policies based on their **substantive scope**, grouping them into four distinct classes. This classification evaluates policy scope rather than operational maturity, and it is evaluated independently of the maturity grid.

Class	Substantive scope	In the report's sample	Share
A	Generic student conduct policy with a brief AI clause added.	4 of 38	approx. 11 percent
B	Dedicated AI policy whose scope is limited to plagiarism and authorship rules.	11 of 38	approx. 29 percent
C	Dedicated AI policy extending into research integrity, validity, and reproducibility.	17 of 38	approx. 45 percent
D	Comprehensive policy addressing AI literacy, core competencies, and thesis examination rules.	6 of 38	approx. 16 percent

Table 11. Policy classification categories by substantive scope, sample counts, and percentage distribution.

Class C represents the most common institutional approach, with Classes C and D together accounting for twenty-three of the thirty-eight universities in the companion report's audit (approximately 60 percent). The remaining 40 percent maintain policies that do not extend beyond basic plagiarism rules.

Policy scope and operational maturity are independent measurements, and both variations occur in practice. An institution with a sophisticated Class C policy may sit at *nascent* on Dimension 3 because it lacks technical procurement controls. Conversely, an institution with a basic Class B policy may achieve an *established* rating on Dimension 3 due to robust central IT security and software management. Both metrics should be recorded without inferring one from the other. Section 6 of Instrument A1 provides eleven straightforward questions to determine an institution's policy class directly from its policy text.

Two additional points are worth noting. First, the practical difference between Class C and Class D involves three specific elements: defining clear competency expectations for students and supervisors, formalizing training pathways, and establishing explicit examination procedures for oral defenses and thesis declarations. These are drafting and workflow updates rather than expensive capital projects, making Class D an achievable goal for any Class C institution. Second, policy classifications vary widely within national jurisdictions (with institutions in the same country sitting simultaneously at Class B and Class D), meaning that choosing peer comparison institutions based on general reputation rather than published policy scope will lead to misleading conclusions.

Part 5 — The modules

The pack contains thirteen instruments organized across four modules. This section outlines the purpose of each instrument, who completes it, the required time commitment, and how it connects to the rest of the toolkit. Appendix A provides this information in a consolidated reference table.

5.1 Module A — the assessment suite

Five instruments that produce, test, interpret, and report the institutional readiness profile.

A1 — Institutional AI-Readiness Maturity Self-Assessment

The primary instrument. Sixty-eight evidence-based indicators evaluating the twenty cells of the grid, alongside eleven questions to determine the Class A to D policy classification. This is the only instrument in the pack that evaluates the complete grid.

Every indicator evaluates practices that an external auditor could verify or refute from documentary records, and each includes an explicit prompt identifying the expected artifact. For example, one indicator asks: *"Does that document publish a task-level demarcation (a named catalogue of research tasks classed as work AI may carry and work AI may not substitute for) rather than a statement of principle alone?"* with the required evidence specified as: *"attach the task table or list with its clause reference, and state how many research tasks it classifies."* The instrument contains no subjective self-ratings.

Who completes it: a cross-functional working group convened by the executive sponsor (see §3.2). **Time:** three to four weeks of document collection, a half-day facilitated workshop, and a two-week window to verify parked items. **Outputs:** twenty cell levels, five dimension levels, five binding axes, the overall binding dimension, a binary Dimension 5 assurance status, a documentary coverage percentage, an assessment verdict, and a policy classification. **Connections:** feeds data directly into A4, where selected operational claims are tested against survey evidence, and is summarized for governance in A5.

The instrument is accompanied by a calculation spreadsheet that records indicator responses and automatically computes rung completions, cell scores, dimension levels, binding axes, the binary Dimension 5 assurance status, and data coverage rates. Printable scoring worksheets are also included in the instrument.

A2 — Researcher and Doctoral-Student Survey

Forty-two closed questions, one short-text item, and two optional open-comment questions across ten sections, examining self-reported research practices, exposure,

knowledge, and context. It asks which AI tools researchers report using, what guidance they have received, what verification checks they report performing, how they handle disclosures, what formal training they completed, what discussions occurred with supervisors, and what data types they report entering into AI platforms. A twenty-four-question short form can be completed in five minutes.

The survey provides conceptually relevant evidence across nine of the twenty cells. It intentionally omits the five policy-axis cells because whether a policy exists is a documentary question answered by A1. A2 evaluates whether those policies actively influence daily student research.

Who completes it: all enrolled doctoral candidates, invited through an attempted institutional census. It can also be extended to post-doctoral and contract research staff if desired. **Time:** ten minutes. **Connections:** feeds directly into Instrument A4, where directly aligned operational items can test selected A1 claims and the remaining items provide context for interpretation and planning. The instrument includes clean text formatted for easy import into standard survey platforms, while administrative analysis keys and crosswalk tables are included in the appendix for the analyst.

A3 — Supervisor Survey

Twenty-six questions across seven sections, with twelve questions linked to related topics in A2. These support direct comparison only where the referent, time period, denominator, and response scale align. Otherwise, the two cohort results are reported side by side. The survey covers supervisory workloads, agreed AI research protocols, faculty knowledge of institutional policies, AI tools used in supervision, actions taken when unverified AI claims are identified, and completed professional development.

Two design elements are important to note. Sections C and G ask supervisors to answer from memory before checking policy references, providing a realistic measure of baseline policy awareness. Additionally, question S1 evaluates whether Instrument G1 (the supervisor-student agreement) is actively utilized in supervision.

Who completes it: all faculty members currently registered as doctoral supervisors, completed anonymously. **Time:** ten to twelve minutes. **Connections:** administered concurrently with A2, feeding directly into the reconciliation analysis in A4.

A4 — Gap-Analysis Worksheet and Action Ladder

Organized into two main sections. **Part 1** provides the analytical framework: it establishes the A1 documentary level, tests directly observable operational criteria against predeclared A2 and A3 evidence, preserves disagreements, applies the reconciliation rules, identifies binding axes and the binding dimension, and executes the Dimension 5 assurance check. It does not average unlike survey items into cell scores. It includes a

twenty-row triangulation table, an institutional reporting profile, and an assurance checklist.

Part 2 contains a sixty-work-package action ladder. Each of the twenty cells has three priority work packages organized around the target transitions from *absent to nascent*, *nascent to established*, and *established to leading*. These are planning packages, not a one-to-one crosswalk to the A1 criteria. Completing one work package is not, by itself, proof of a level transition. Before commissioning it, the institution names the exact A1 indicators it intends to rescore. After any package, it must conduct a fresh A1 rescore of every applicable indicator at the target rung and below for the named cell. The cell advances only when all are Yes, and the dimension advances only when all four cells qualify at the target level. Partial-rung work remains supporting rather than level-completing. Each package has a designated owner, a required evidence artifact, and an estimated timeframe. A representative entry reads: *"Build and publish a register of the AI tools actually used in research, assembled from expenditure data, tenant sign-in logs, and a short return from each school and college. Record for each entry: the tool class, the owner, the tier, the data classes it touches, and where the data resides."* (Assigned to the Chief Information Officer and Faculty Associate Deans for Research, supported by a published register containing those five fields, with an estimated timeframe of six to ten weeks).

The instrument also provides ninety-day and twelve-month implementation roadmaps for the standard institutional profile, customized sequences for policy-heavy or decentralized universities, and a summary of common implementation pitfalls.

Who completes it: the designated benchmarking owner in collaboration with the graduate school, research integrity office, and IT leadership. **Time:** three to four hours for Part 1 with completed assessments in hand, and fifteen minutes per selected action item in Part 2. **Connections:** synthesizes data from A1, A2, and A3, and generates the executive data for A5.

A5 — Board / Council Reporting One-Pager

A single-page executive summary designed for university governing bodies. It presents the institutional profile in a clean five-row table displaying each dimension's score, its binding axis, progress since the previous reporting cycle, and a concise summary of the primary bottleneck. It also highlights the overall binding dimension, outlines three concrete resource requests, reviews regulatory compliance timelines, and provides a brief note explaining why maturity is not reduced to a single score. A dedicated progress column supports consistent reporting across regular governance cycles.

Who completes it: the designated Dimension 5 owner, countersigned by the reporting executive. **Time:** thirty minutes once A4 is completed, as this document summarizes the reconciled profile rather than generating new data.

5.2 Module B — the briefings

Five concise briefings tailored for specific leadership roles, each addressing three core questions: why this issue matters to your role, which specific decisions you uniquely own, and what questions you should ask your team this month. Each briefing is designed to be read independently without requiring prior study of the companion report. They explicitly reference the relevant grid cells and cross-reference corresponding sections in the report.

These briefings serve as an effective entry point for institutional engagement. In practice, universities often circulate B1 to executive leadership, B2 to faculty deans, B3 to graduate school leadership, B4 to research integrity committees, and B5 to graduate student associations, initiating the broader assessment process from the resulting discussions.

Briefing	Audience	Read time	Core decisions addressed
B1	Vice-Presidents for Research, Deputy Vice-Chancellors, Provosts	10 min	Deciding whether AI in research is governed under research integrity or student conduct, appointing the benchmarking owner, establishing enterprise software gates for unpublished research data, approving foundational budgets, and establishing public compliance positions for upcoming regulatory audits.
B2	Faculty Deans, Associate Deans for Research, Department Chairs	15 min	Defining task-level rules for academic disciplines, establishing faculty-specific task schedules, routing department software through procurement gates, integrating AI competencies into local curricula, and evaluating departmental maturity against university-wide baselines.
B3	Deans of Graduate Studies, Directors of Doctoral Colleges	15 min	Establishing graduate school policies with explicit examination rules, integrating the six core competencies into doctoral training, delivering supervisor professional development, and requiring supervisor-student AI agreements.
B4	Research Integrity and Graduate Studies Committees	12 min	Defining genuine research integrity violations in AI usage, establishing adjudication frameworks for AI misconduct cases, setting practical disclosure standards, and explaining why automated detection tools cannot replace human verification.
B5	Graduate Student Associations and Student Leaders	15 min	Identifying six reasonable expectations graduate students should have regarding AI governance, highlighting key academic risks facing students, defining student advocacy priorities, and using published maturity standards to evaluate university support.

Table 12. Module B briefing documents with target audiences, estimated reading times, and core governance decisions.

Briefing B4 opens directly with the core challenge facing integrity committees: explaining why procedures built for traditional plagiarism often fail to catch the data fabrication and citation errors most commonly generated by AI systems. Briefing B5 is written from the

student perspective, noting that since universities are not legally required to run these assessments, having published, transparent maturity rubrics gives student representatives a valuable tool for constructive advocacy.

5.3 Module G — the agreement

G1 — Supervisor–Student AI-Use Agreement

The flagship operational instrument, designed to embed institutional standards directly into everyday research supervision. While many university policies state high-level principles regarding human oversight, very few clarify where that line falls for a specific doctoral project in a specific discipline. Central policy cannot anticipate the nuances of every research methodology. That determination is best made collaboratively by the student and supervisor who understand the project.

Instrument G1 is completed in a single thirty-minute discussion, followed by brief ten-minute check-ins during scheduled milestone reviews. At its core is a **task schedule** containing twenty-one standard research tasks organized across six phases (literature search, study design, data collection and analysis, manuscript drafting, figures and visualizations, and final thesis defense), alongside blank rows for discipline-specific tasks. For each task, the student and supervisor select one of three protocols: *AI-assisted, with disclosure*, *Human-only*, or *Negotiated per instance*. Certain fundamental tasks are designated as permanently human-only by the framework: supervisor approval of analysis chapters, examiner evaluation at the oral defense, authorship determinations, and peer review assessments, as these represent points where human scholars must take personal responsibility for academic claims.

The agreement also includes six operational sections: confirming which institutional policies outrank the agreement, establishing a ten-field prompt and output logging protocol, signing an explicit list of confidential data that must never be uploaded to external tools, scheduling regular milestone reviews, defining a dispute resolution process, and clarifying thesis submission requirements. It concludes with a sign-off section and a fully worked example.

Who completes it: the doctoral student and principal supervisor together, with co-supervisors countersigning. **Outputs:** a signed research protocol, a designated log repository, and an agreed review schedule (three concrete artifacts suitable for graduate school audits). **Evaluates:** D1–process , D1–people , D2–process , D2–people , D3–systems , and D4–people .

Two key design choices are worth emphasizing. First, the *Negotiated per instance* option allows students to use AI for brainstorming or exploring alternatives (such as generating counter-arguments or reviewing code structure) while ensuring that the final analytical decisions remain entirely their own. The framework prohibits delegating final scientific

decisions to software, but encourages using tools to support human critical thinking. Second, G1 is structured to protect both the student and the university. Agreed, logged, and verified AI usage has no automatic retroactive effect under a later policy change, subject to applicable law, ethics, contractual, confidentiality, funder, publisher, or other binding requirements. The agreement gives the student clear documentation to support their thesis defense.

Instrument G1 was authored independently based on the task taxonomy in Appendix G of the companion report. Prior published work, notably the University of New Hampshire's *Collaboration Agreement for Use of Generative AI in a Research Project Template* (April 2026), is acknowledged as the closest prior example, though G1 adopts an independent structure and operational phrasing.

5.4 Module P — the policy library

P1 — Model Institutional AI-in-Research Policy

A complete, adoptable institutional policy template comprising fifteen clauses and four operational schedules, drafted so university legal counsel and research offices can adapt it into formal governance without starting from scratch. Its structure covers: institutional purpose, scope and application, definitions, core principles, human-in-the-loop requirements, permitted and restricted uses by research activity, disclosure requirements, data security classifications and tool tiers, procurement controls, supervision and doctoral training expectations, thesis examination procedures, publication and grant submission standards, administrative roles and responsibilities, proportionate misconduct responses, and regular policy review schedules.

The template is drafted to satisfy Class D standards, synthesizing the actual practices of the six Class D universities identified in the companion report's audit. If an institution prefers to target a Class C scope, it can adopt the template while omitting the specific competency and examiner clauses. This is an entirely legitimate strategic choice, but it should be made explicitly rather than through drafting oversights.

Every clause includes two annotations: a **drafting note** explaining the rationale behind the language and suggesting local adaptations, and a **framework crosswalk** identifying the corresponding grid cells, companion report sections, and specific audit artifacts required to demonstrate that the clause is actively functioning. Jurisdictional notes highlight where specific phrasing should be tailored to local legislation. These annotations are intended for the drafting committee and are removed prior to publication.

An important note on maturity levels: adopting P1 supplies evidence toward D1-policy through D5-policy at the *established* rung. It does not establish a cell or dimension. A fresh A1 assessment determines cell levels, and a rescore across all four cells determines each dimension's level. Reaching the *leading* level requires published evaluation data, verified

outcome metrics, and external peer benchmarking, as evaluated in Instruments A1, A4, and A5.

P2 — AI-Use Disclosure Standard

A standardized institutional policy defining when AI usage must be disclosed, what information the disclosure must contain, and where disclosures belong across various research outputs (including dissertations, journal articles, grant proposals, conference presentations, datasets, software code, and technical reports), complete with customizable statement templates.

Its underlying approach addresses two common misconceptions about disclosure. First, disclosure exists solely to ensure research reproducibility and provenance. **Disclosure is not an admission of error or a disciplinary penalty**, and the standard states this explicitly. When disclosure is perceived as risky, compliance drops dramatically. For example, a study in the *BMJ* identified an initial disclosure rate of only 5.7 percent across 25,114 submissions to 49 medical journals after introducing a mandatory disclosure field, during a period when AI tools were already widely used for literature reviews, editing, and coding. Researchers will not provide transparent disclosures if the process is cumbersome or perceived as punitive.

Second, the standard emphasizes that disclosure alone does not validate research findings. While disclosure provides necessary transparency, true scientific validity depends on rigorous verification against primary sources, which remains an independent research obligation.

Who adopts it: the relevant university academic board or research committee as a binding standard. **Time:** forty-five minutes to review local adaptations, ninety minutes of committee discussion to adopt, and three minutes for a researcher to prepare a standard disclosure statement. **Connections:** represents the operational disclosure clause of P1 published as a standalone standard. While G1 establishes what research uses are permitted, P2 defines how those uses are formally recorded.

5.5 How the modules connect

The toolkit follows an integrated implementation sequence across one academic quarter:

1. **Leadership briefings are circulated.** Briefing B1 goes to executive leadership, while B2 through B5 are distributed concurrently to deans, graduate school leaders, integrity committees, and student representatives.
2. **A benchmarking lead is appointed** and the institutional mandate is published. This supplies evidence toward D5—people → nascent and takes less than an hour of executive time. It does not move the cell or dimension by itself. A fresh A1 rescore determines the cell, and Dimension 5 advances only when all four D5 cells qualify at the target level.

3. **Documentation is collected and A1 is scored** in a facilitated workshop, while student (A2) and supervisor (A3) surveys are administered in parallel.
4. **Instrument A4 reconciles the three evidence streams**, establishing recorded cell levels, identifying binding constraints, executing the Dimension 5 assurance check, and generating a prioritized twelve-month action plan with assigned owners.
5. **Findings are reported to governance bodies** using the A5 executive template, and the university determines what information to publish.
6. **Policy templates P1 and P2 provide the formal text** to address identified gaps on the policy axis.
7. **Agreement G1 is deployed across doctoral programs**, establishing clear supervisor-student protocols that are subsequently measured in future survey cycles through question S1 in A3.

Universities operating on compressed timelines can complete steps 1, 2, 3, and 5 initially, deferring the remaining steps. However, institutions facing near-term regulatory deadlines should prioritize building their software registers and completing impact assessments early, as scoping an impact assessment requires first having an accurate inventory of the AI tools currently used across research projects. Building that inventory typically requires six to ten weeks of administrative effort.

Part 6 — A worked example: Northgate University

Northgate University is an illustrative case study. Its institutional profile is designed to reflect typical sector challenges rather than an idealized institution, and all data points are internally consistent with the scoring logic in the companion calculation workbook.

6.1 The institution

Northgate University is a comprehensive, research-intensive institution comprising six faculties, including a medical school, with 1,430 enrolled doctoral candidates and 947 registered doctoral supervisors. It co-operates a joint doctoral program with an EU partner university and utilizes an automated screening tool to process doctoral scholarship applications, which places at least one of its operational systems within the high-risk education classification governed by Article 27 of the EU AI Act from 2 August 2026.

Northgate previously published *Generative AI in Research and Research Training* (2025), a standalone guidance document issued by its central Research Committee, distinct from its

general student conduct rules. Substantive research integrity expectations are set out in its Research Integrity Framework 2025. The university completed the readiness assessment during the third quarter using an eleven-person working group facilitated by the University Secretary's office, fielding surveys A2 and A3 during the same period.

Survey participation reached **511 of 1,430 doctoral students (36 percent)** and **268 of 947 supervisors (28 percent)**. Both cohorts exceeded Northgate's pre-established response threshold of 25 percent overall with at least ten responses per faculty, making the results eligible for the predeclared criterion-specific checks. Eligibility did not make every survey item level-changing.

6.2 The twenty cells, scored

Across the sixty-eight indicators in Instrument A1, Northgate recorded **25 Yes, 3 Partial, 33 No, and 7 unanswered** responses at the two-week resolution deadline. The unanswered indicators contributed no points and blocked confirmation of their rungs, but they were not recoded as No. Their affected cells were marked unresolved at the relevant rung.

A1 documentary coverage: 61 of 68 indicators = about 90 percent. Documentary verdict: Final.

Cell	Nascent rung	Established rung	Leading rung	Cell level	Binding evidence gap
D1-policy	met	met	unresolved	Established	Leading-level evaluation evidence unresolved at the deadline.
D1-people	met	met	unresolved	Established	Leading-level publication and improvement evidence unresolved at the deadline.
D1-systems	met	not met	unresolved	Nascent	Thesis declaration is unformatted text rather than structured task categories. Leading evidence unresolved.
D1-process	met	met	unresolved	Established	Leading-level audit evidence unresolved at the deadline.
D2-policy	met	met	unresolved	Established	Leading-level compliance and review evidence unresolved at the deadline.
D2-people	met	not met	unresolved	Nascent	Four-mode rules distributed only as an unmonitored PDF guide. Leading evidence unresolved.
D2-systems	met	not met	unresolved	Nascent	AI disclosures stored only as unstructured text within thesis files.

Cell	Nascent rung	Established rung	Leading rung	Cell level	Binding evidence gap
					Leading evidence unresolved.
D2-process	met	not met	not met	Nascent	No formal verification of disclosures during thesis submission.
D3-policy	met	not met	not met	Nascent	Procurement policy evaluates only two of six properties and lacks the required exception protocol.
D3-people	met	not met	not met	Nascent	No software evaluation panel, and privacy and integrity officers are omitted from reviews.
D3-systems	not met	not met	not met	Absent	No centralized inventory of AI tools used in research.
D3-process	met	not met	not met	Nascent	Article 27 impact assessment for the admissions screening tool not completed.
D4-policy	met	in progress	not met	Nascent	Policy specifies only four of the six core competencies as learning outcomes.
D4-people	not met	not met	not met	Absent	No institutional training offering or designated

Cell	Nascent rung	Established rung	Leading rung	Cell level	Binding evidence gap
					academic lead is in place.
D4-systems	met	not met	not met	Nascent	Supported delivery lacks versioned curriculum content, aligned assessments, and retrievable completion records.
D4-process	met	not met	not met	Nascent	No scheduled review cycle for research methods curricula.
D5-policy	met	not met	not met	Nascent	Assessment not conducted against a recognized published framework.
D5-people	met	not met	not met	Nascent	No assigned ownership map across the twenty grid cells.
D5-systems	not met	not met	not met	Absent	Compliance records gathered manually from disparate offices on request.
D5-process	met	not met	not met	Nascent	Assessment reports not formally tabled at council or faculty boards.

Table 13. Scoring results across twenty cells showing rung attainment, overall cell level, and evidence gaps.

The seven unresolved indicators affected D1-policy, D1-people, D1-systems, D1-process, D2-policy, D2-people, and D2-systems. Each is shown as unresolved at the relevant leading rung. Where a documented No already prevented an earlier rung, the unresolved leading item did not change the demonstrated lower cell level.

Three cells produced particularly important diagnostic findings:

D3-policy is classified as *not met* at the established level. Northgate's software procurement guidelines evaluate only two of the six required technical properties: data residency and log auditability. They do not evaluate citation verification, uncertainty reporting, reproducibility at fixed model versions, or local open-source deployment options, and they lack the required exception protocol. Under A1, that documented No makes the rung *not met*. Identifying the missing criteria in the evidence register still provides a clear, manageable drafting task.

D4-policy is similarly classified as *in progress*. Northgate's doctoral regulations list four of the six competencies as learning outcomes (citation verification, model specification, sycophancy detection, and structured failure reporting), but omit prompt sensitivity management and model heterogeneity in adversarial review. This also results in a *Partial* rating, identifying another straightforward policy update.

D1-systems illustrates a very common disconnect. Northgate's electronic thesis submission portal has included a generative AI declaration field since 2025. However, it is an unstructured free-text box that does not require candidates to categorize usage against the task schedule published in the university's own Research Integrity Framework. The policy schedule and the technical form both exist, but they have never been connected.

6.3 The dimensional profile

Dimension	policy	people	systems	process	Level (minimum)	Binding axis
D1 Human-in-the-loop discipline	Established	Established	Nascent	Established	Nascent	systems
D2 Responsible use in practice	Established	Nascent	Nascent	Nascent	Nascent	people, systems, process
D3 Tooling that promotes responsible use	Nascent	Nascent	Absent	Nascent	Absent	systems
D4 AI-literate humans	Nascent	Absent	Nascent	Nascent	Absent	people
D5 Institutional benchmarking grid	Nascent	Nascent	Absent	Nascent	Absent	systems

Table 14. Dimensional profile showing axis ratings, minimum maturity levels, and identified binding axes.

Binding dimension: D3 (Tooling that promotes responsible use). Three dimensions share the minimum score of *absent*: D3, D4, and D5. Applying the framework's sequencing rules resolves the tie in favor of the lowest-numbered dimension (D3). This prioritization makes practical sense: Northgate cannot effectively design its AI literacy curriculum (D4) or finalize its compliance benchmarking (D5) without first knowing which AI tools its researchers are using.

Dimension 5 assurance check: not satisfied. Because D5 sits at *absent*, well below the required *established* baseline, all four of Northgate's documented *established* cells (D1-policy, D1-people, D1-process, and D2-policy) must be recorded as **unassured**. Northgate may use these scores internally, but it cannot publish them or quote them in accreditation reviews until central governance reaches *established*.

This finding was particularly valuable for Northgate's leadership. The university has done solid work on Dimension 1: it published a detailed task schedule, integrated task boundaries into supervisor training (achieving a 78 percent completion rate among active supervisors), updated external examiner guidelines, and established clear procedures for

thesis evaluations. However, it cannot readily substantiate these achievements to an external auditor because it lacks a centralized records register. The underlying capability exists, but formal assurance does not.

Analyzing the institutional profile. Northgate exhibits the most common pattern across higher education: formal policy running well ahead of administrative execution. This is especially clear in Dimension 2, where policy is *established* while people, systems, and processes remain *nascent*. In this environment, effective practice depends entirely on individual departmental initiative, and the university cannot identify where compliance is strong or weak. Drafting another policy document will not resolve this bottleneck. Northgate already excels at policy drafting, and adding more documentation would create an illusion of progress without improving operational practice.

6.4 Triangulation: what the surveys changed

Northgate predeclared criterion-specific survey checks for directly observable claims at the *established* level. Policy-existence findings remained documentary questions. The table below is a screening summary. Where it does not identify an exact criterion-specific reconciliation record, the survey result is treated as contextual and cannot support an alignment claim or change the documented level.

Cell	Documented level	A2 student evidence	A3 supervisor evidence	Survey reading	Reconciled level	Gap type
D1-policy	Established	<i>not surveyed</i>	76% policy awareness	Not tested against survey evidence	Established	Documentary-only / not tested
D1-people	Established	31% recipient report	68% supervisor report	Not directly aligned with the A1 supervisor-development and attendance criterion	Established	Transmission diagnostic
D1-process	Established	71%	79%	Context only; the criterion-specific record is not shown	Established	Documentary-only / not tested
D2-policy	Established	<i>not surveyed</i>	72% policy awareness	Not tested against survey evidence	Established	Documentary-only / not tested

Table 15. Reconciliation of documented cell levels with criterion-specific student and supervisor evidence.

No cell changed as a result of the criterion-specific checks. The D1-people documentary level remained *established* because the student item could not directly test whether the supervisor-development program included the required content and maintained attendance records.

The 31 percent student result and 68 percent supervisor result still identify a clear transmission diagnostic. They concern whether guidance reached its intended recipients, not whether the A1 D1-people supervisor-development and attendance criterion was satisfied. Northgate therefore reported the two anonymous cohorts side by side. If the institution uses this comparison to change a level in a future cycle, it must first tie it to a separate, predeclared recipient-facing process criterion with the same referent and time period.

Northgate's reconciled profile following triangulation:

Dimension	Level	Binding axis
D1 Human-in-the-loop discipline	Nascent	systems
D2 Responsible use in practice	Nascent	people, systems, process
D3 Tooling that promotes responsible use	Absent	systems
D4 AI-literate humans	Absent	people
D5 Institutional benchmarking grid	Absent	systems

Table 16. Reconciled dimensional profile showing maturity levels and binding axes after survey triangulation.

Binding dimension: D3 (unchanged). Documentary verdict: Final (90 percent A1 coverage). Assurance status: unassured (due to D5 remaining below *established*).

Triangulation revealed two strategic insights that were not apparent from the document review alone. First, the *systems* axis is the binding constraint across three of the five dimensions, making it the university's primary technical priority for resource allocation. Second, the 31 percent and 68 percent results identify a transmission diagnostic for a future recipient-facing process check, while the documented binding axis in Dimension 1 remains *systems* because the thesis submission portal fails to record structured task data.

6.5 Class A–D placement

Evaluated directly from policy documentation across eleven standardized diagnostic questions:

Gate	Diagnostic question	Answer
1	Does a dedicated AI policy exist, distinct from general student conduct rules?	Yes (<i>Generative AI in Research and Research Training</i> , 2025).
1	Does the policy apply specifically to doctoral research rather than coursework alone?	Yes (Scope clause 1.2).
1	Was it issued by central university governance or the graduate school?	Yes (Research Committee).
2	Does the policy address research validity and reproducibility, rather than attribution alone?	Yes.
2	Does it explicitly address citation fabrication and require verification against primary sources?	Yes.
2	Does it address AI as an analytical validator, distinct from writing assistance?	Yes.
2	Does it require disclosure across research outputs, rather than assessed coursework alone?	Yes.
3	Does it define explicit competency standards for graduate students?	Yes (four competencies specified).
3	Does it define explicit competency standards for supervisors?	Partial (phrased only as general aspirations).
3	Does it reference a formal curriculum or required training pathway?	No (relies entirely on voluntary library resources).
3	Does it define explicit examination rules for oral defenses and thesis declarations?	Partial (provides general defense expectations, but lacks mandatory verification requirements or examiner report fields).

Table 17. Diagnostic gate questions and evaluation answers used for policy classification placement.

Classification: Class C (Gate 3 requirements not fully satisfied).

Northgate sits within the most common national policy category (Class C, representing 45 percent of the companion report's audit sample) while three of its five operational dimensions remain *absent*. This divergence illustrates why policy scope and operational maturity must be evaluated separately. Northgate has developed a respectable policy on paper, but lacks the administrative systems to execute it effectively.

This classification also identifies Northgate's most cost-effective path to policy leadership. Moving from Class C to Class D requires three specific updates: defining precise competency expectations for supervisors, establishing a formal training pathway, and adding structured AI verification fields to thesis examination forms. These are administrative and drafting tasks that can be completed within standard governance cycles.

6.6 The three work packages Northgate should complete first

Three specific cells hold entire dimensions at *absent*: D3-systems, D4-people, and D5-systems. These represent Northgate's immediate operational priorities and can be pursued concurrently.

Each item below is a priority work package toward the named target rung, not proof of a completed transition. Northgate must conduct a fresh A1 rescore of every applicable indicator at that rung and below after the work is completed. The named cell advances only when all are Yes, and its dimension advances only when all four cells qualify at the target level. Partial-rung work remains supporting.

Work package 1 (toward D5-systems → nascent): establish a central evidence register.

Construct a centralized register documenting the specific artifact, location, designated owner, and review date for each of the twenty grid cells. *Owner*: Director of Research Integrity. *Required artifact*: an active register with all twenty rows completed. *Estimated timeframe*: four to eight weeks. *Why first*: this is the most straightforward task and provides the underlying evidence repository needed for Northgate eventually to clear the Dimension 5 assurance check. Moving D5-systems to *nascent* does not clear that check, which requires all four D5 cells to reach *established*.

Work package 2 (toward D3-systems → nascent): construct the research AI tool inventory.

Build and maintain an inventory of AI tools used across research activities, compiled from IT licensing records, single-sign-on logs, and departmental returns. For each tool, document its functional category, designated owner, licensing tier (consumer or enterprise), data sensitivity classification, and data residency location. *Owner*: Chief Information Officer, in collaboration with Faculty Associate Deans for Research. *Required artifact*: a published software register containing all five fields, with an established update schedule. *Estimated timeframe*: six to ten weeks. *Why second*: this represents the binding axis of the primary binding dimension (D3) and sits on the critical path for regulatory

compliance. Scoping the EU AI Act Article 27 impact assessment for the scholarship screening tool requires first having an accurate inventory of deployed systems.

Work package 3 (toward D4-people → nascent): appoint an academic lead for AI research training. Appoint a designated academic lead for AI research competencies, supported by a formal time allocation within the institutional workload model. An unresourced voluntary role will produce limited practical results. *Owner:* Vice-President for Research, in consultation with the Dean of Graduate Studies. *Required artifact:* an approved role description, formal workload allocation, and an official appointment record. *Estimated timeframe:* six to ten weeks. *Why third:* while Dimension 4 is an important long-term priority, formal curriculum changes operate on multi-year academic cycles and cannot be verified without the evidence register established in Rung 1. Appointing the academic lead establishes leadership for that work, but it does not advance D4-people while the institutional training offering required by the same nascent criterion is not present.

What Northgate should avoid this quarter. It should not draft additional policy documents, as its policy axis is already well developed. It should not procure new AI software tools until its procurement standards are finalized in Dimensions 1 and 2. Finally, it should not attempt to jump directly from *absent* to *established* in a single step, as establishing working systems requires first creating the foundational nascent artifacts.

6.7 What the profile looks like twelve weeks later

Completing these three foundational work packages requires roughly ten weeks of administrative coordination, a single academic appointment, and no capital expenditure. The table shows the result only after Northgate rescored the applicable A1 indicators.

Dimension	Before	After	Binding axis after
D1 Human-in-the-loop discipline	Nascent	Nascent	systems
D2 Responsible use in practice	Nascent	Nascent	people, systems, process
D3 Tooling that promotes responsible use	Absent	Nascent	all four axes tied
D4 AI-literate humans	Absent	Absent	people
D5 Institutional benchmarking grid	Absent	Nascent	all four axes tied

Table 18. Comparison of institutional dimensional maturity levels and binding axes before the work packages and after the A1 rescore.

Two dimensions advance after the rescore. D3 and D5 move from *absent* to *nascent* because the work packages satisfy the applicable cell criteria and the remaining cells in each dimension were already at *nascent*. D4 remains *absent* because appointing the academic lead alone does not satisfy D4-people while the training offering is missing. The new evidence register can still support the next cycle of work in D1, including connecting the task schedule to the thesis submission portal and examining whether supervisory guidance reaches students through a separately defined recipient-facing process criterion.

The Dimension 5 assurance check will continue to apply until all four D5 cells reach *established*. Clearing it will require completing the twenty-cell ownership map, formally adopting the framework within institutional policy, establishing the required systems evidence, and presenting regular assessment reports to the university council and faculty boards. On Northgate's governance calendar, this represents a realistic nine-to-twelve-month work plan toward assured *established* and *leading* claims in Dimensions 1 to 4.

Part 7 — Adapting the pack

7.1 The invitation

I encourage you to adapt these instruments to fit your institution's specific needs. That flexibility is the explicit purpose of releasing this pack under an open license.

The toolkit is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0). You are free to copy, share, modify, translate, incorporate into internal policies, use within commercial training or consulting, and redistribute every instrument in this pack, provided you include appropriate attribution. You do not need to seek permission or notify Instats, though I welcome feedback and case studies at support@instats.org.

This permissive license is intentional because these instruments are most effective when tailored to local contexts. For example, an institution with an existing thesis declaration should integrate these AI disclosure fields directly into that form rather than creating a separate document. A computational biology department might expand G1's task schedule with rows for protein structure prediction, while a history department might add rows for archival transcriptions. Similarly, universities operating in jurisdictions without specific AI legislation should adapt the regulatory indicators to reflect their local standards. The core capability this framework evaluates is the institutional discipline of defining clear expectations for your own researchers, rather than adhering rigidly to the templates in this pack.

7.2 What attribution looks like in practice

Under the CC BY 4.0 license, attribution simply means ensuring that readers can identify the original source and recognize where modifications have been made. In practice, this involves three basic elements:

1. **Acknowledge the original source, author, and license.**
2. **Provide a link to the original materials** where feasible.
3. **Indicate that modifications were made**, if applicable.

A standard attribution statement in a document footer or colophon satisfies these requirements:

Adapted from The Institutional AI Readiness Pack by Michael J. Zyphur (Instats), licensed under CC BY 4.0, available at instats.org/publications/the-institutional-ai-readiness-pack. Modified for use at [INSTITUTION NAME]. The task schedule was customized for local discipline requirements, and log retention schedules and escalation pathways were established in accordance with institutional policies.

For an instrument used without significant changes, a shorter statement is sufficient:

Source: The Institutional AI Readiness Pack by Michael J. Zyphur (Instats), CC BY 4.0, instats.org/publications/the-institutional-ai-readiness-pack.

For translated materials:

Translated from The Institutional AI Readiness Pack by Michael J. Zyphur (Instats), CC BY 4.0. Translation prepared by [TRANSLATOR / INSTITUTION NAME] and not reviewed by the original author.

Every document in this pack includes a standard attribution footer. Retaining that footer is the simplest way to satisfy licensing requirements. If you remove the footer, simply include an equivalent attribution note.

Two practical points: attribution does not require repeating full license text inside individual policy clauses (a single note in your policy history or administrative colophon is

entirely sufficient), and attribution does not imply that Instats has formally reviewed or endorsed your institution's specific policies.

7.3 Translation and local adaptation

I strongly encourage translations and regional adaptations. The framework evaluates standards applicable across fifteen countries, but the initial instruments are published in English with common Anglophone administrative titles. Adapting these materials for other languages and governance structures is essential for broader adoption.

Four guidelines for translators and regional adapters:

Maintain consistent terminology within your adaptation. The core vocabulary (including the five dimension titles, four operational axes, four maturity levels, four AI-use modes, six tool properties, and six core competencies) appears across the thirteen instruments and Start Here. Use consistent terms throughout your translation, and include a brief glossary mapping your terms to the original English definitions so readers can easily reference the companion report.

Preserve the standard cell codes. Identifiers such as D1-policy, D3-systems, and D5-process are language-neutral codes that link survey items, policy clauses, and action items directly to the same grid cell. Translate the descriptive text, but retain the cell identifiers.

Adapt regulatory references to your local legal context rather than omitting them. This pack references specific European and Australian regulations because they carry firm compliance deadlines. If those specific laws do not apply to your institution, substitute the applicable national laws, privacy standards, or voluntary sector frameworks in your jurisdiction. The indicators evaluate whether your institution has assigned ownership and established timelines for its legal compliance, regardless of which specific statutes apply.

Use your institution's actual administrative titles. The instruments use standard titles (such as Vice-President for Research, Dean of Graduate Studies, or Research Integrity Officer) alongside common international equivalents. Substitute your university's actual role titles before distributing briefings or surveys, as materials addressed to unfamiliar job titles are less likely to engage busy colleagues.

7.4 Two things to avoid

Do not calculate a single, aggregated institutional score. Creating an overall summary score undermines the primary diagnostic value of the assessment, which is to identify the specific operational bottlenecks holding an institution back. If senior leadership requests a

high-level summary, provide the five dimension scores and the binding dimension, which fits clearly on a single line and points directly to actionable priorities.

Do not present modified instruments as the original, unedited framework. Label adaptations clearly so reviewers, peer institutions, and student representatives know which version they are evaluating. Transparent benchmarking depends on everyone using the same stated criteria.

Appendix A — Instrument catalogue

ID	Instrument	Purpose	Who completes or reads it	Length	Time
A1	Institutional AI-Readiness Maturity Self-Assessment	Scores the twenty cells using documentary evidence and determines dimension levels, binding constraints, and policy classification (Classes A to D).	Cross-functional working group convened by the Vice-President for Research or Provost.	15–20pp + workbook	3–4 weeks for document collection, a half-day scoring workshop, and a 2-week resolution period.
A2	Researcher and Doctoral-Student Survey	Collects student evidence on daily research practices, training exposure, knowledge, and context across 42 closed questions, 1 short-text item, and 2 optional open-	All enrolled doctoral candidates, invited through an attempted institutional census.	6–8pp	10 minutes (5 minutes for the 24-question short form).

ID	Instrument	Purpose	Who completes or reads it	Length	Time
		comment questions.			
A3	Supervisor Survey	Gathers supervisor reporting on training, supervision practices, and policy awareness across 26 questions, including 12 linked to related A2 topics.	All registered doctoral supervisors, completed anonymously.	4–6pp	10–12 minutes.
A4	Gap-Analysis Worksheet and Action Ladder	Reconciles the three evidence streams, identifies binding constraints, executes validation checks, and selects from sixty supporting work packages.	Designated benchmarking lead, in consultation with graduate school, integrity, and IT leads.	10–14pp	Part 1: 3–4 hours. Part 2: 15 minutes per selected action item.
A5	Board / Council Reporting One-Pager	Summarizes the reconciled institutional profile, binding constraints, and resource requests for university governing bodies without aggregating to a single score.	Designated Dimension 5 owner, countersigned by the reporting executive.	1–2pp	30 minutes once A4 is completed.

ID	Instrument	Purpose	Who completes or reads it	Length	Time
B1	Briefing: executive research leadership	Reviews five core decisions owned by executive leadership, immediate diagnostic questions, and resource requirements.	Vice-Presidents for Research, Deputy Vice-Chancellors, Provosts.	2–3pp	10 minutes.
B2	Briefing: deans and associate deans of research	Addresses faculty-level governance, local research workflows, and managing disciplinary differences effectively.	Deans, Associate Deans for Research, Department Chairs.	2–3pp	15 minutes.
B3	Briefing: graduate school leadership	Identifies key touchpoints where AI intersects doctoral supervision, research training curricula, and thesis examinations.	Deans of Graduate Studies, Directors of Doctoral Colleges.	2–3pp	15 minutes.
B4	Briefing: research-integrity and graduate studies committees	Defines genuine research integrity risks in AI usage, outlines a four-part case adjudication model, and reviews	Members of research integrity boards and graduate studies committees.	2–3pp	12 minutes.

ID	Instrument	Purpose	Who completes or reads it	Length	Time
		detection limits.			
B5	Briefing: graduate-student associations	Outlines six reasonable student expectations, evaluates key academic risks, and provides constructive advocacy tools.	Graduate student associations and student governance representatives.	2–3pp	15 minutes.
G1	Supervisor–Student AI-Use Agreement	A structured protocol for students and supervisors to agree upon task-by-task AI usage boundaries, logging rules, and review schedules.	Doctoral candidates and principal supervisors collaboratively, with co-supervisors countersigning.	3–4pp	30 minutes for initial completion and 10 minutes per milestone review.
P1	Model Institutional AI-in-Research Policy	A comprehensive Class D policy template comprising 15 clauses and 4 schedules, complete with legal and framework annotations.	Research office and legal counsel, in consultation with graduate school and integrity leads.	8–12pp	2–3 drafting sessions plus one standard institutional review cycle.
P2	AI-Use Disclosure Standard	A standardized institutional rule governing disclosure across all research output types,	Formally adopted by academic board and implemented by researchers,	4–6pp	45 minutes to review and 90 minutes of committee discussion to adopt.

ID	Instrument	Purpose	Who completes or reads it	Length	Time
		complete with statement templates.	supervisors, and students.		
—	The Institutional AI Readiness Pack (this document)	The comprehensive citable overview: restating the framework, methodology, interpretation rules, worked case study, and catalogue.	University leaders and working groups planning and conducting institutional assessments.	50–70pp	45 minutes to read, with Parts 3, 4, and 6 serving as active working sections.

Table 19. Catalog of assessment pack instruments listing purpose, target audience, document length, and completion time.

Appendix B — TEQSA crosswalk

Australia's Tertiary Education Quality and Standards Agency (TEQSA) has issued some of the most comprehensive sector guidance on generative AI in higher education, and Australian universities should use this pack in conjunction with those materials. This appendix maps the pack to TEQSA's frameworks. While drafted for Australian institutions, this alignment logic applies equally to other jurisdictions with comparable national quality standards.

Distinguishing between the two TEQSA toolkits. TEQSA has released two distinct publications:

- *Gen AI strategies for Australian higher education: Emerging practice* (November 2024) is the broader sector-wide toolkit developed following an information request to all Australian higher education providers. It focuses primarily on academic integrity, coursework assessment, and teaching and learning, **introducing the Process / People / Practice framework.**

- ***Gen AI strategies for research training: Emerging practice*** (June 2025, 26 pages) focuses specifically on doctoral education, organized into four operational sections with binary compliance checklists and institutional case studies.

Framework alignment and distinct contributions. The Process / People / Practice model in TEQSA's November 2024 toolkit aligns closely with the four governance axes in the companion report (policy, people, systems, and process). This independent structural convergence reinforces the validity of both models. The companion report's framework builds on this foundation by adding an explicit **policy** axis and distinguishing between **systems** (underlying technical infrastructure and data registers) and **process** (supervisory and administrative workflows), recognizing that technical and administrative failures require different solutions and owners.

How the toolkits work together: TEQSA defines what Australian institutions should achieve. This pack provides the practical tools to implement those practices and assess institutional maturity. The two frameworks are complementary. TEQSA's research training toolkit provides an excellent checklist and case study collection, but does not include scoring rubrics, maturity grids, survey instruments, or policy templates. This pack provides those operational instruments, enabling institutions to establish documentary levels, test selected operational claims against respondent evidence, and manage structured improvement plans.

The crosswalk

TEQSA research training checklist section (June 2025)	Checklist expectations	Corresponding pack instruments
1. Induction, guidance and training	Inducting candidates and supervisors into AI expectations, providing accessible guidance, and delivering structured training.	G1 (structured supervisor-student agreement) · P1 cl. 10 (supervisory and training obligations) · A1 indicators for D4-people, D4-process, and D1-process · A2 §6–§7 and A3 §F (verifying training delivery) · A4 action rungs for D4-people and D1-process · B3 (graduate school briefing).
2. The research process	Governing AI usage in research conduct, including data protection, methodological rigor, analysis, and primary verification.	P1 cl. 5, 6, and 8 (human oversight, permitted uses, data classification, tool tiers) · G1 §4B–§4C and §6 (task schedules and restricted data lists) · A1 D1 (all axes), D2-process, D3-systems, D3-process · A2 §2, §4, §8 (student tool usage, checks, and data inputs) · A4 rungs for D1-policy and D3-systems.
3. Assessment and thesis examination	Aligning examination practices with AI realities, including candidate declarations, examiner guidelines, and oral defense expectations.	P1 cl. 11 (thesis examination standards) · G1 §9 (submission and defense protocols) · P2 §4 (disclosure locations by output type) · A1 D1-systems, D1-process, D2-process (declaration fields, examiner packs, report forms) · A3 §E · B4 (integrity committee briefing) · A4 action rungs for D1-systems and D2-process.
4. Publications and grant applications	Ensuring transparent disclosure in publications and grant proposals aligned with publisher and research funder requirements.	P2 in full (disclosure thresholds, required elements, output locations, and customizable statement templates) · P1 cl. 7 and cl. 12 · A1 D2-policy, D2-systems · A4 action rungs for D2-policy and D2-systems.

Table 20. Mapping of regulatory research training checklist sections and expectations to corresponding pack instruments.

Using the two together

Australian universities can efficiently utilize both frameworks within the same review cycle. The most effective approach is to review the TEQSA checklist first to identify which broad areas the university currently addresses, complete Instrument A1 to establish documentary maturity levels across the twenty cells, and then administer surveys A2 and A3 to test directly observable operational claims and describe daily practice. The TEQSA checklist confirms broad policy coverage, while this pack evaluates operational maturity and assurance.

Australian institutions should also note that the Australian Privacy Act 1988 automated-decision-making transparency requirements commence on 10 December 2026 for private universities and the Australian National University. This statutory obligation is distinct from TEQSA guidelines and spans `D5-policy` (a documented plan and owner) and `D5-process` (deadline recognition and formal governance review).

Appendix C — AI-assistance disclosure for this pack

This disclosure documents the use of AI assistance in preparing this pack, following the transparency standards established in Appendix D of the companion report.

AI assistance. This report used assistance from three model families while I drafted and edited the pack. Naming them precisely is the standard P2 asks institutions to meet, so this disclosure does the same.

Tool	Version and setting	Used for
Anthropic Claude	Opus 5, at extra and maximum reasoning effort	Initial drafting of the instruments, structural design of the assessment, and orchestration of the work
OpenAI Codex	GPT-5.6-sol, at high and extra-high reasoning effort	Adversarial review, claim checking against sources with web search enabled, and verification of cross-document consistency
Google Gemini	3.7 Flash, at high reasoning effort	Copy-editing, voice and readability revision, and review of the rendered page layouts

Table 21. Disclosure of artificial intelligence tools, version configurations, and specific drafting applications used.

Draft materials were developed across the thirteen instruments and Start Here, and aligned with the controlled terminology and scoring logic defined in the companion report.

Scope of AI tool usage. Tools were used to generate initial drafts from detailed structural outlines, to verify that indicators and cell codes matched across the thirteen instruments and Start Here, to confirm cross-references to the companion report, to check the internal calculations in the case study, and to copy-edit. A later pass revised the prose of every document for readability, removing constructions that read as machine-written rather than human-written.

Activities excluded from AI delegation. No empirical claims, interpretations of literature, statistical figures, compliance dates, or citations were generated by AI tools without direct verification against the companion report and primary evidence dossiers. I authored the core framework architecture, maturity criteria, reconciliation methodology, licensing decisions, and strategic analysis directly.

Author responsibility. I take full responsibility for all factual statements, source interpretations, calculations, analytical conclusions, and final phrasing across this document and the thirteen accompanying instruments.

Conflict-of-interest statement. My professional role at Instats involves research methods training. I developed this framework and toolkit independently of Instats' commercial operations and publish them as an open contribution to the international higher education community. The pack is licensed under CC BY 4.0 so institutions can adopt and modify these instruments freely without any commercial engagement with Instats.

Appendix D — How to cite this pack, and the sources it draws on

Citing the pack

To cite the complete toolkit:

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

To cite an individual instrument:

Zyphur, M. J. (2026). *A1 — Institutional AI-Readiness Maturity Self-Assessment*. In *The Institutional AI Readiness Pack*. Instats Policy Series.

To cite the companion report, containing the full evidence base and theoretical framework:

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

If you are citing the overall theoretical model (the five dimensions, four axes, four maturity levels, and twenty cells), cite the companion report. If you are citing a specific diagnostic instrument, indicator, or implementation action item, cite this pack.

What this pack draws on

The substantive foundations of this pack derive from the companion report, which provides the formal definitions, scoring criteria, and empirical evidence summarized here. Key mappings include:

Element of the pack	Report source
The five dimensions and core definitions	Part 2, §2.1–§2.5
The four operational axes	§2.5
The four maturity levels and criteria	§4.1
The twenty-cell maturity grid	Appendix A
The cumulative scoring rule	Appendix A (<i>leading</i> criteria build on <i>established</i>)
The dimension-as-minimum scoring rule	§4.1 (conjunctive definition of <i>established</i>)
The rationale against single summary scores	§4.3
The Dimension 5 validation check	§4.1
The Class A to D policy taxonomy	§1.3, §3.2
The research task schedule (behind G1 and D1)	Appendix G
The AI tool classifications (behind D3 and P1)	Appendix C
The four AI-use modes	§2.2
The six technical tool properties	§2.3
The six core competencies	§2.4
Fixed-date regulatory requirements	§2.5, §4.2
The forward-looking standard for Dimension 5	§4.2
Leadership briefing recommendations (B1 to B5)	§5.1–§5.5
Regional adaptation principles	Appendix F

Table 22. Mapping of assessment pack components to foundational sections in the companion report.

The companion report's empirical findings rest on eight primary evidence dossiers covering national research funder policies, university AI regulations, competency models,

publisher and journal standards, AI tool taxonomies, international data privacy legislation, research reproducibility literature, and automated adversarial review studies. Every key empirical finding in the report is linked to primary sources with archived snapshot dates.

Prior published work. The University of New Hampshire's *Collaboration Agreement for Use of Generative AI in a Research Project Template* (April 2026) is acknowledged as an early institutional example of a supervisor-student agreement. Instrument G1 was authored independently using the task schedule in Appendix G of the companion report. The two documents maintain independent structures and are licensed separately.

Sector guidance. When referencing guidance from national regulators or sector bodies (such as the TEQSA analysis in Appendix B), this pack cites the official titles and publication dates. Institutions using external guidance for legal or accreditation compliance should always consult the original regulatory publications directly.

Appendix E — Crosswalk to the Responsible AI framework

Every section in this document maps directly to corresponding sections in the companion report, *Responsible AI in Academic Research: A Competency Framework for Research Training*.

Section of this document	Companion report section
Executive summary: baseline findings and policy audit	Executive Summary; §1.1; §1.3; §3.2
Executive summary: Dimension 5 leadership standard	§4.2
Executive summary: rationale against single scores	§4.3
Executive summary: regulatory compliance timelines	§2.5
1.1 The ladder rung nobody has reached	§4.2; §2.5
1.2 The problem an institution cannot currently answer	Executive Summary; §1.2; §1.3; Appendix C (Class 11 detection tools)
1.3 What this pack is, and what it is not	§4.1; §4.3; Appendix F
2.1 The five dimensions	Part 2 Introduction; §2.1; §2.2; §2.3; §2.4; §2.5
2.1 Footnote on Dimension 2 naming	§2.2 heading; §4.1 table; Appendix A row 2
2.2 The four axes	§2.5
2.3 The four levels	§4.1; Appendix A
2.4 The twenty-cell grid	§4.1; Part 2 summary table; Appendix A
2.5 The four AI-use modes	§2.2
2.5 The six tool properties	§2.3
2.5 The six competencies	§2.4
3.1 Triangulation methodology	§4.1 (operational definitions); Appendix A
3.2 Working group composition	§2.5 (people axis); §5.1–§5.4
3.3 Resource requirements and timelines	§4.1 (rapid review methods); §5.1
3.3 Defining the unit of assessment	§4.3; §5.2
3.4 Evidence rules and verification standards	§4.1; §2.3 (evaluating vendor claims)

Section of this document	Companion report section
3.5 Common implementation pitfalls	§4.1; §4.3; §2.3
4.1 The cumulative scoring rule	Appendix A (cumulative criteria)
4.2 The dimension-as-minimum rule	§4.1; §2.5
4.3 Identifying the binding axis	§2.5; §4.3
4.3 Four common institutional profile patterns	§4.3 (case study analyses)
4.4 Identifying the binding dimension	Part 2 Introduction (precondition sequence); §4.3
4.5 The Dimension 5 validation check	§4.1
4.6 Data coverage, Final and Provisional status	§4.1 (reviewable schedules); Appendix A
4.7 Rationale against single summary scores	§4.3
4.8 Class A to D policy classification	§1.3; §3.2
5.1 Module A (assessment instruments)	Part 4; Appendix A (survey methodology)
5.2 Module B (leadership briefings)	§5.1; §5.2; §5.3; §5.4; §5.5
5.3 Module G (supervisor-student agreement)	§2.1; §2.2; Appendix G; §5.3
5.4 Module P (policy templates and disclosure standards)	§1.3; Part 2; §2.2; §3.1; §3.3; Appendix C; Appendix G
5.5 Implementation workflow	§5.1; §4.2
6.1–6.3 Worked case study: scoring and profiles	§4.1; §4.3; Appendix A
6.4 Triangulation analysis	§4.1
6.5 Policy classification evaluation	§1.3
6.6–6.7 Action prioritization and twelve-week progress	§5.1; §5.3; §2.3; §2.5

Section of this document	Companion report section
7.1–7.4 Adapting and translating the toolkit	Appendix F (fidelity, currency, attribution)
Appendix A: Instrument catalogue	Part 4; Part 5
Appendix B: TEQSA sector crosswalk	§3.4
Appendix C: AI assistance disclosure	Appendix D (transparency model)
Appendix D: Citation guidelines and primary sources	Appendix D; Appendix E

Table 23. Comprehensive crosswalk mapping sections of this document to corresponding companion report sections.

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>

License. The pack and its instruments are licensed under [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). You may adapt them for institutional use with attribution.

1. The companion report refers to this dimension using two slightly different titles. The heading in §2.2 reads "Responsible use across the four AI-use modes," while the Part 4 summary table, Appendix A, and the project repository use "Responsible use in practice." In this pack, I use **Responsible use in practice** consistently across all instruments and crosswalks so readers always know which dimension is being discussed. ↩