A European University Alliance

invest

The Competitive University for a Sustainable Society

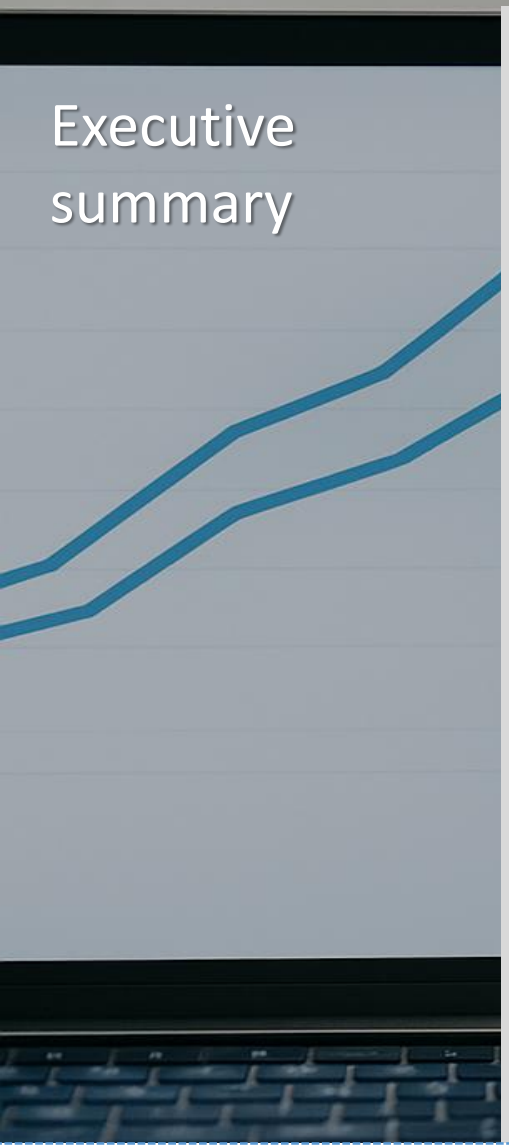
Towards Responsible AI-Supported PhD Supervision

A Policy Perspective on the Tripartite Mentoring Model

POLICY BRIEF

Co-funded by the European Union

Executive summary



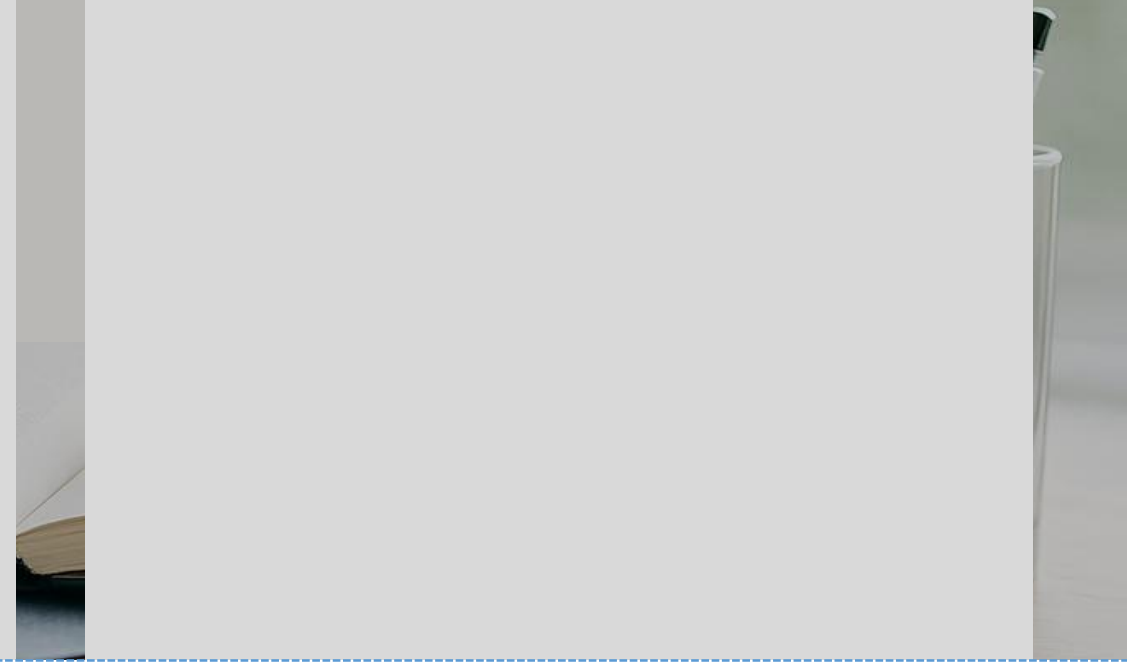
The rapid advancement of generative AI is beginning to reshape doctoral education, raising urgent questions about how tools such as ChatGPT can be integrated into PhD supervision without weakening academic integrity or the human core of mentorship. This policy brief synthesizes evidence from the first blinded, expert-led feasibility study of ChatGPT-4 for doctoral guidance, conducted on a real PhD supervision case developed within the INVEST European University Alliance by the University of Thessaly partner.

Moving beyond perception-based research, the study generated AI guidance under four prompting strategies and subjected the outputs to blinded review by five external academics using a validated six-dimension rubric (factual accuracy, academic depth, methodological coherence, contextual relevance, critical thinking, practical applicability). Expert consensus on quality rankings was strong (Kendall's $W=0.648$, $p<0.05$), confirming the reliability of the evaluation.

Findings show that AI output quality is highly prompt-dependent: topic-specific, structured prompts were rated appropriate by 100% of experts (5/5), while generic keyword prompting produced markedly weaker guidance (40%, 2/5). This evidence supports the Tripartite Mentoring Model, which frames doctoral supervision as a collaborative triad between student, supervisor, and AI assistant, governed by 7 principles including complementarity, AI literacy, ethical oversight, and critical mediation.

Policy implications point to the need for institutional guidelines on responsible AI use, targeted capacity building for supervisors and candidates in prompt design and critical evaluation, and phased piloting of AI-augmented supervision within quality-assured frameworks.

Read the full study in the International Journal of Doctoral Studies: https://www.informingscience.org/Publications/5579

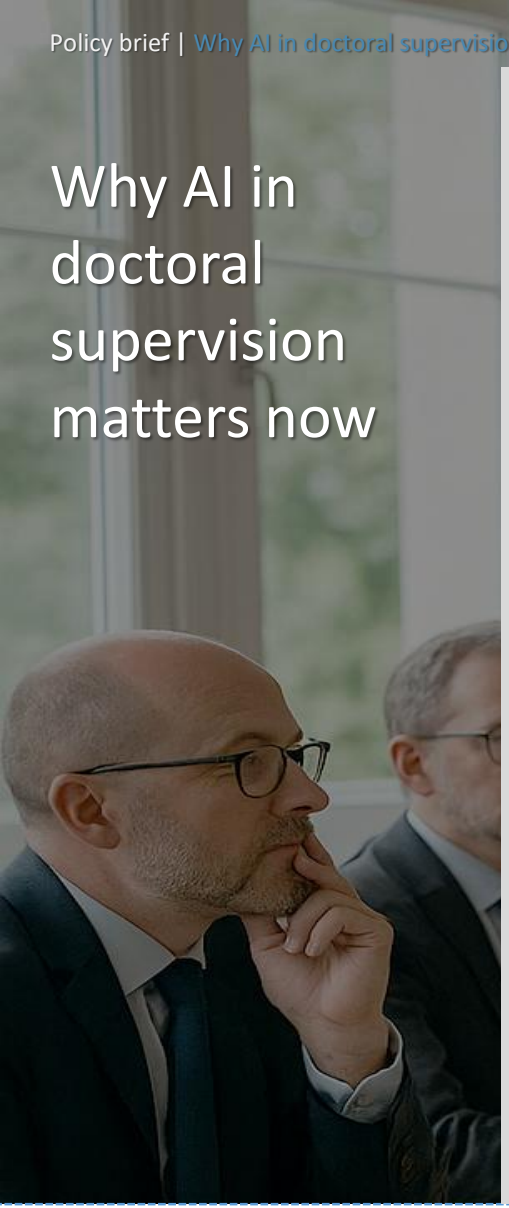


Prompting drives quality: structured prompts make ChatGPT PhD-useful

Tripartite model: AI supports - humans lead supervision

Responsible adoption needs policy + training: AI literacy, attribution rules, and data-safe practices are essential

Why AI in doctoral supervision matters now



Generative AI is no longer peripheral to doctoral work; it is already being used informally by PhD candidates for literature scanning, idea generation, methodological clarifications, and drafting support. At the same time, doctoral supervision across Europe is under growing pressure: research topics are increasingly interdisciplinary, supervisory workloads are rising, and candidates often experience delayed or fragmented feedback, especially at early stages when uncertainty is highest. These conditions create a real demand for scalable forms of academic support that can complement—rather than dilute—the supervisory relationship.

AI assistants like ChatGPT can help address this gap by providing rapid, accessible, and context-sensitive guidance when prompts are well-structured, supporting candidates with concept clarification, research framing, outlining, and exploratory methodology suggestions. In the INVEST alliance context, where cross-institutional supervision, mobility, and joint doctoral pathways are strategic priorities, such tools could strengthen continuity of support across partners and reduce friction in the early shaping of research directions.

However, the opportunity comes with clear risks. The study emphasizes that GenAI outputs may include factual inaccuracies, shallow reasoning, or misalignment with disciplinary standards, especially when prompts are vague or under-specified. Without AI literacy and critical mediation, candidates may over-trust the tool, weaken their independent scholarly judgment, or drift into questionable authorship and integrity practices. These concerns are not hypothetical; they are intrinsic to how large language models operate and are amplified when guidance is taken at face value rather than evaluated academically.

This creates a policy window: universities and doctoral schools need to move from informal, uneven use to structured, ethically governed integration. AI should be framed as a bounded cognitive tool within a human-led mentoring ecosystem, supported by training, transparency, and clear institutional rules. The evidence in this paper shows that the quality and safety of AI support depend strongly on how it is used—making proactive governance essential before practices become entrenched by default.

GenAI is already influencing PhD work; supervision needs proactive, structured integration, not ad-hoc adoption

AI's value in doctoral guidance is real but conditional: without AI literacy and critical oversight, risks to rigor and integrity grow

What the Evidence Shows

A feasibility study was conducted within the INVEST alliance using a real PhD supervision case developed by the University of Thessaly partner in the domain of disaster risk management. The purpose was to test whether ChatGPT-4 can provide doctoral-level research guidance that external academics consider appropriate, and under what conditions.

To ensure a robust assessment beyond user perceptions, ChatGPT outputs were generated under **four prompting strategies** that varied in structure and contextual richness: a naive summary prompt, keyword-enhanced prompts (keywords selected by supervisors vs. keywords suggested by ChatGPT), and a topic-specific prompt enriched with concepts derived from Structural Topic Modeling of relevant literature. Each output was independently reviewed by **five blinded external experts** from complementary fields (AI in education, disaster risk management, data science/predictive analytics, and doctoral-level academic writing/research methods).

Experts rated appropriateness using a 5-point scale across six quality dimensions: factual accuracy, academic depth, contextual fit, critical thinking, methodological rigor, and practical applicability. Inter-rater agreement on quality rankings **was strong and statistically significant** (Kendall's $W = 0.648$, $p < 0.05$), indicating reliable convergence of expert judgments.

Results show a clear and policy-relevant pattern: prompt structure is a decisive determinant of doctoral guidance quality. The naive and topic-specific (STM-enriched) prompts were rated appropriate by 100% of experts (5/5), while keyword-enhanced prompts performed unevenly: supervisor-selected keywords reached 80% appropriateness (4/5), but ChatGPT-generated keywords fell sharply to 40% (2/5). Importantly, longer AI responses were not necessarily better; output length increased with richer prompts, but quality depended on specificity and reasoning structure, not verbosity.

A secondary analysis tested whether ChatGPT can support doctoral decision-making when asked to advise along distinct research pathways already identified by the supervisory team. Three pathway-specific prompts were evaluated—Geospatial Intelligence & Remote Sensing, Digital Twin technologies, and Semantic Web approaches—and were judged mostly appropriate (80%, 80%, and 100%, respectively). This suggests that when supervision provides bounded direction and context, GenAI can contribute meaningfully to exploring methods and structuring early research choices—while still requiring human validation.

Blinded experts confirmed that ChatGPT-4 can offer PhD-level guidance, but only when prompts are structured and context-rich

Prompt strategy produced a large quality gap (100% vs 40% appropriateness), making prompt literacy a core policy lever

THE TRIPARTITE MENTORING MODEL

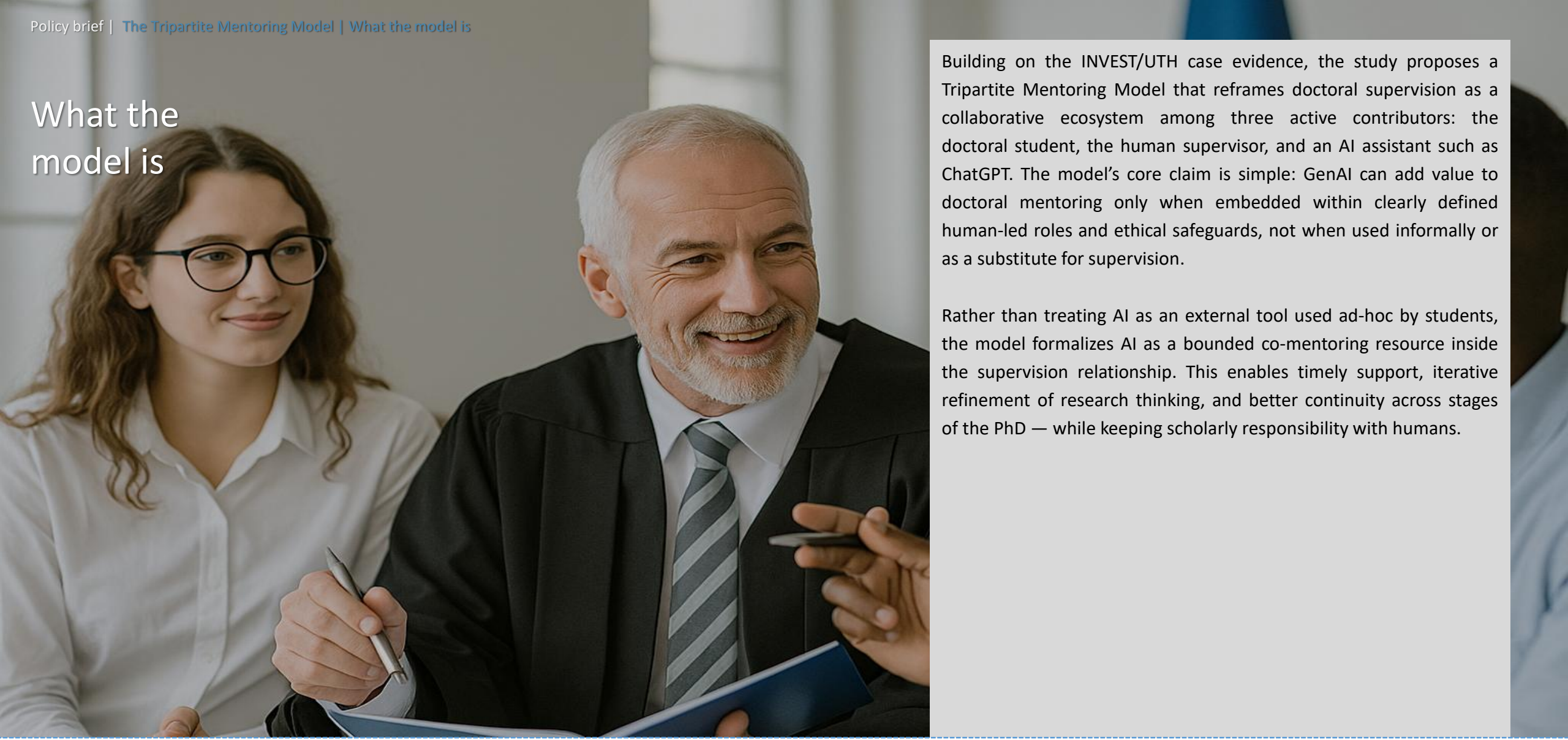
In this section

What the model is

Roles within the supervision triad

Seven governance principles for responsible implementation

What the model is



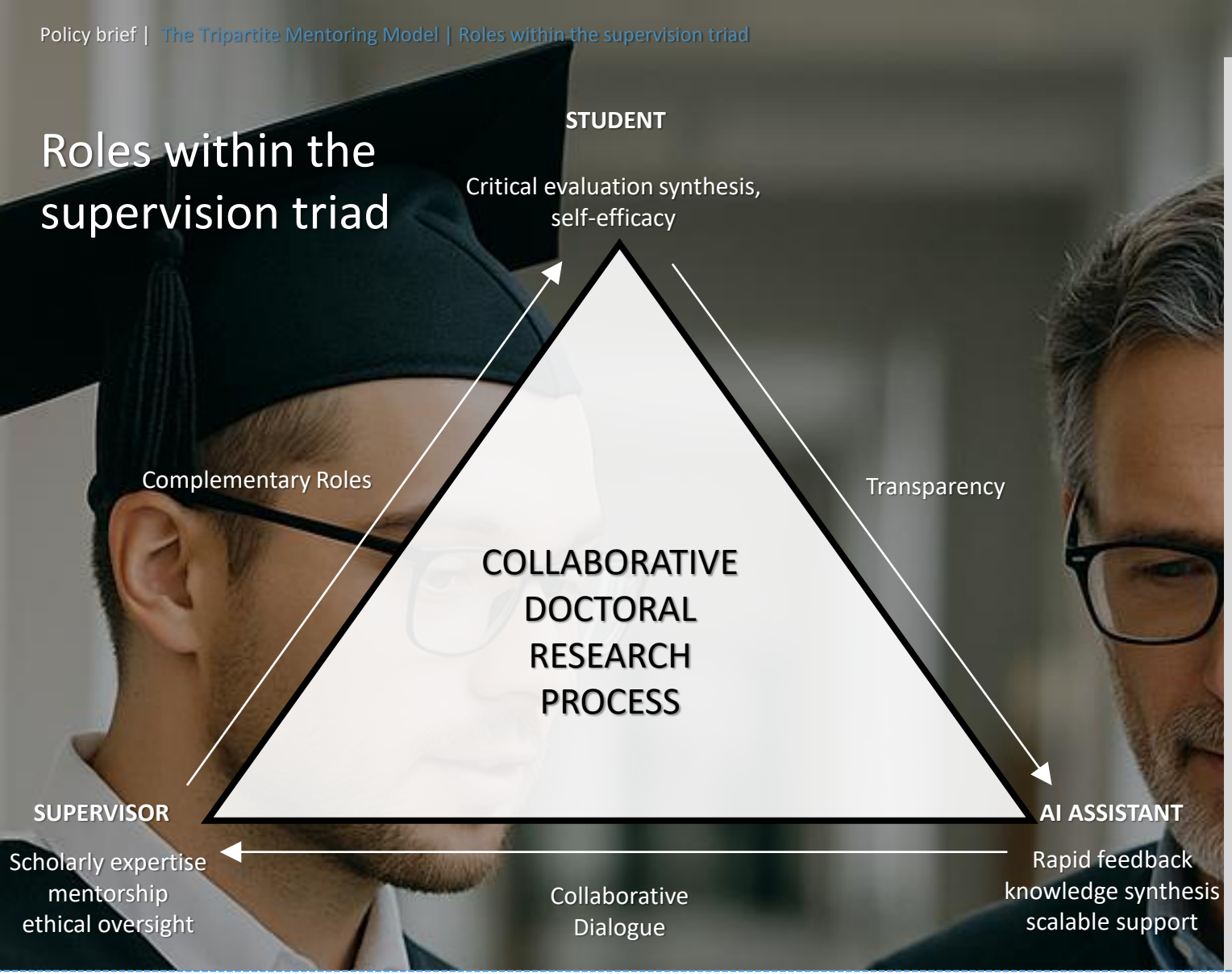
Building on the INVEST/UTH case evidence, the study proposes a Tripartite Mentoring Model that reframes doctoral supervision as a collaborative ecosystem among three active contributors: the doctoral student, the human supervisor, and an AI assistant such as ChatGPT. The model's core claim is simple: GenAI can add value to doctoral mentoring only when embedded within clearly defined human-led roles and ethical safeguards, not when used informally or as a substitute for supervision.

Rather than treating AI as an external tool used ad-hoc by students, the model formalizes AI as a bounded co-mentoring resource inside the supervision relationship. This enables timely support, iterative refinement of research thinking, and better continuity across stages of the PhD — while keeping scholarly responsibility with humans.

The Tripartite Mentoring Model embeds AI as a bounded support partner, with human supervision and student agency remaining central

Seven governance principles translate AI-assisted supervision into a scalable, ethical institutional practice

Roles within the supervision triad



The Tripartite Mentoring Model redefines doctoral supervision as a shared, dynamic process among three actors:

- **Supervisor (human-led authority):** provides domain expertise, scholarly judgment, and quality control. The supervisor validates and contextualizes any AI-supported guidance, and safeguards academic standards and integrity.
- **Doctoral student (agentic researcher):** remains the driver of research decisions and authorship. The student uses AI outputs as material for reflection, critically evaluates them, and synthesizes them with supervisory feedback.
- **AI assistant (bounded cognitive support):** supplies rapid, structured guidance when prompted with adequate context. It supports ideation, clarification, structuring, and exploration of alternatives, but does not make final decisions or own intellectual credit.

This structured collaboration fosters a balance of creativity, guidance, and accountability, supporting transparency and pedagogical integrity. Importantly, it frames AI not as a replacement, but as a complementary tool within a human-led academic ecosystem.

AI complements, not replaces, the supervisor

Seven governance principles for responsible implementation

The model is operationalized through seven interdependent principles that define how the triad should function in practice:

1. Complementary roles and defined responsibilities

Each actor contributes distinct value: supervisor judgment, AI's rapid context-aware feedback, and the student's critical synthesis.

2. Collaborative dialogue and decision-making

Supervision becomes an iterative feedback loop where AI suggestions are discussed, tested, and refined jointly by student and supervisor.

3. Ethical governance and academic integrity

AI use must follow explicit rules on responsible use, fairness, privacy, and protection against over-reliance or bias, with human oversight as the integrity anchor.

4. Transparency of AI involvement

AI contributions should be openly acknowledged within supervision and in scholarly outputs, preventing covert use and authorship ambiguity.

5. AI literacy and capacity building

Students and supervisors need shared competence in prompt design, interpretation, and critical checking, because guidance quality depends heavily on prompting.

6. Personalization and flexibility

AI support should adapt to the doctoral stage and student needs, scaling up or down while human mentorship remains the guiding authority.

7. Ongoing evaluation and accountability

AI-augmented supervision should be monitored through rubric-based review, milestone tracking, and feedback from both students and supervisors, enabling continuous improvement.

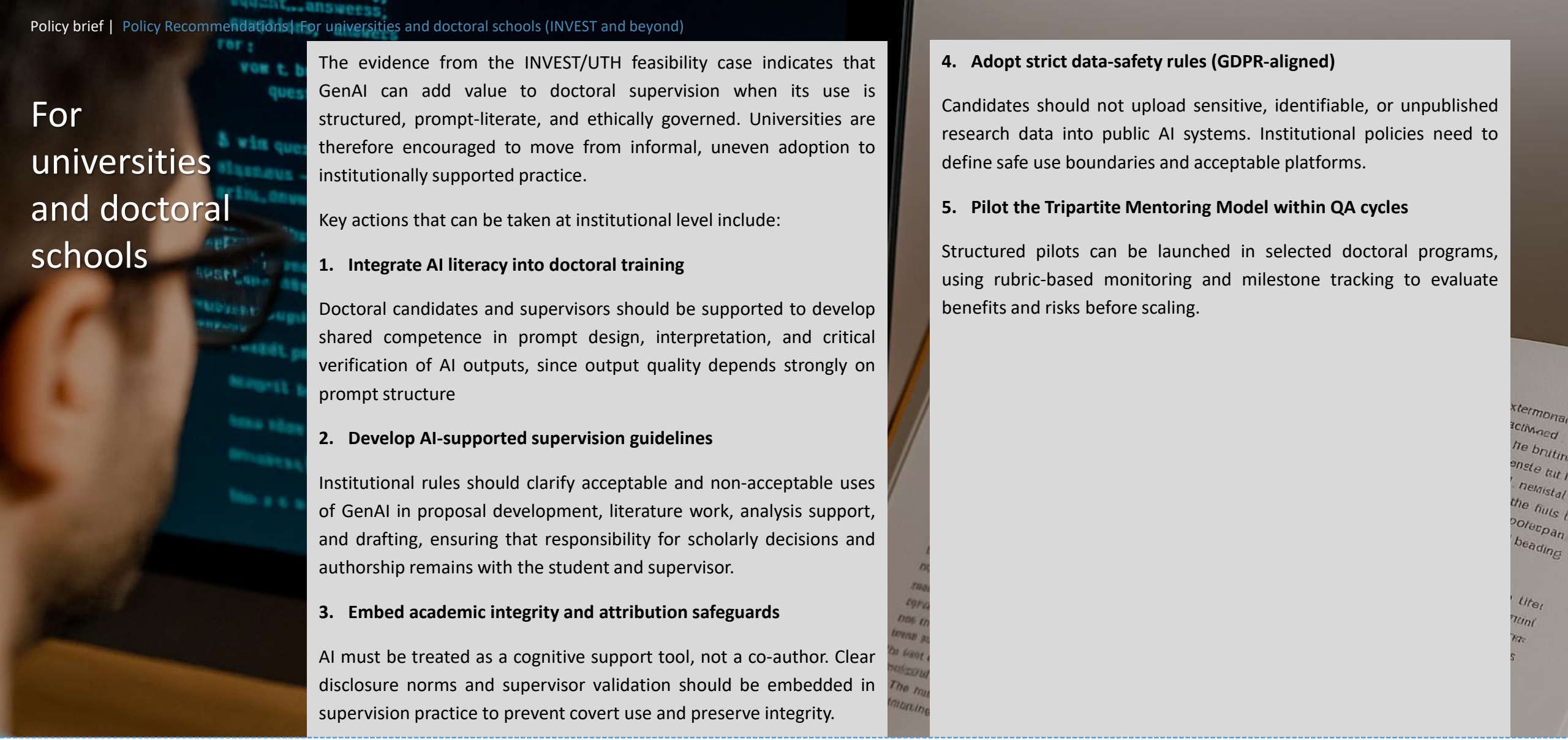
Together, these principles ensure that AI strengthens doctoral **supervision without eroding agency, rigor, or ethics**, and they provide universities with a practical governance blueprint for scaling responsible GenAI use across programs.

Key principles include promoting AI literacy, ensuring transparency in prompting, and upholding academic integrity

The framework is designed to be adaptable to different disciplinary norms and institutional contexts

POLICY RECOMMENDATIONS

In this section



For universities and doctoral schools

The evidence from the INVEST/UTH feasibility case indicates that GenAI can add value to doctoral supervision when its use is structured, prompt-literate, and ethically governed. Universities are therefore encouraged to move from informal, uneven adoption to institutionally supported practice.

Key actions that can be taken at institutional level include:

1. **Integrate AI literacy into doctoral training**

Doctoral candidates and supervisors should be supported to develop shared competence in prompt design, interpretation, and critical verification of AI outputs, since output quality depends strongly on prompt structure

2. **Develop AI-supported supervision guidelines**

Institutional rules should clarify acceptable and non-acceptable uses of GenAI in proposal development, literature work, analysis support, and drafting, ensuring that responsibility for scholarly decisions and authorship remains with the student and supervisor.

3. **Embed academic integrity and attribution safeguards**

AI must be treated as a cognitive support tool, not a co-author. Clear disclosure norms and supervisor validation should be embedded in supervision practice to prevent covert use and preserve integrity.

4. **Adopt strict data-safety rules (GDPR-aligned)**

Candidates should not upload sensitive, identifiable, or unpublished research data into public AI systems. Institutional policies need to define safe use boundaries and acceptable platforms.

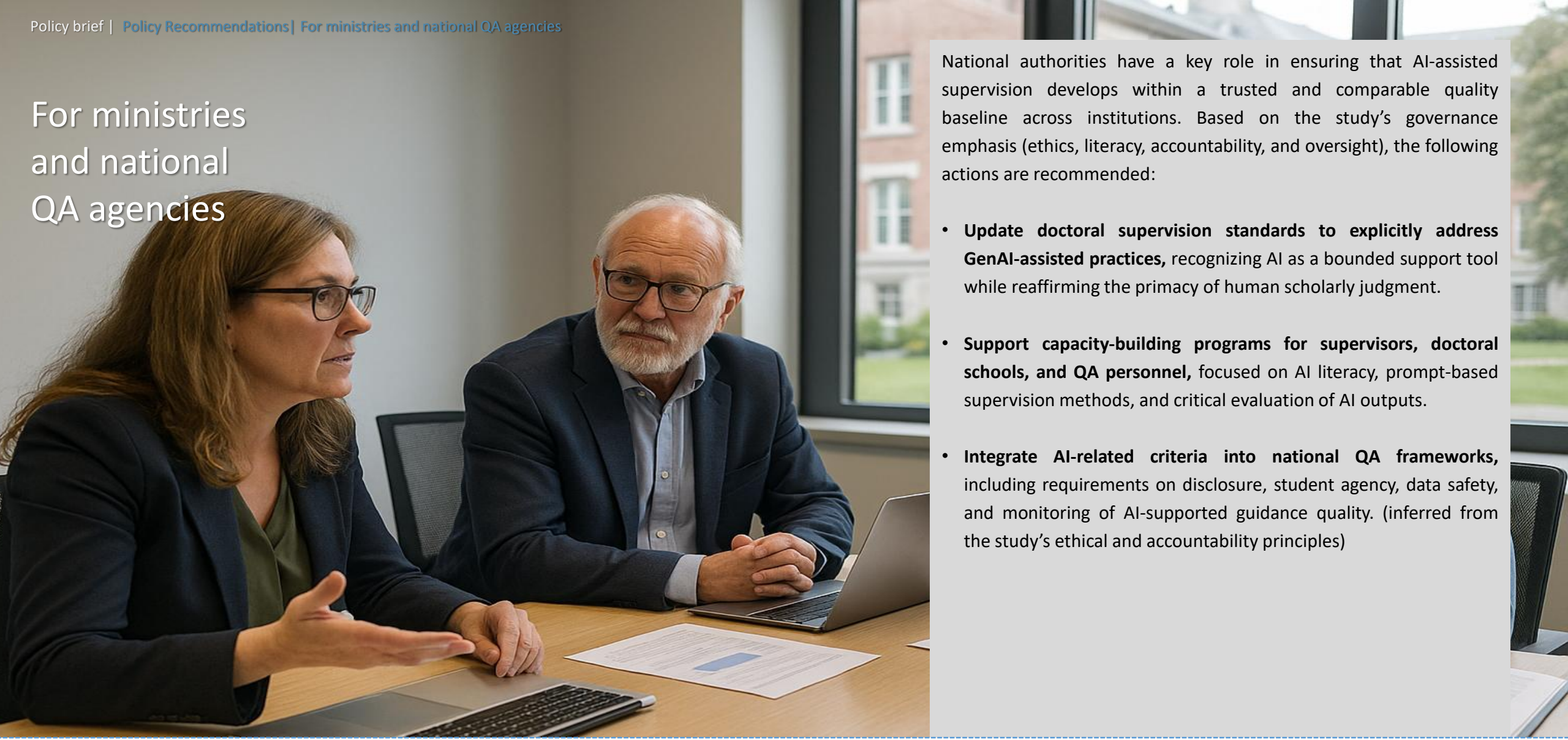
5. **Pilot the Tripartite Mentoring Model within QA cycles**

Structured pilots can be launched in selected doctoral programs, using rubric-based monitoring and milestone tracking to evaluate benefits and risks before scaling.

Universities should institutionalize AI literacy, integrity rules, and GDPR-safe practice before GenAI becomes supervision-by-default

The Tripartite Mentoring Model can be piloted within Universities as a scalable, QA-aligned blueprint for human-led AI-augmented supervision

For ministries and national QA agencies



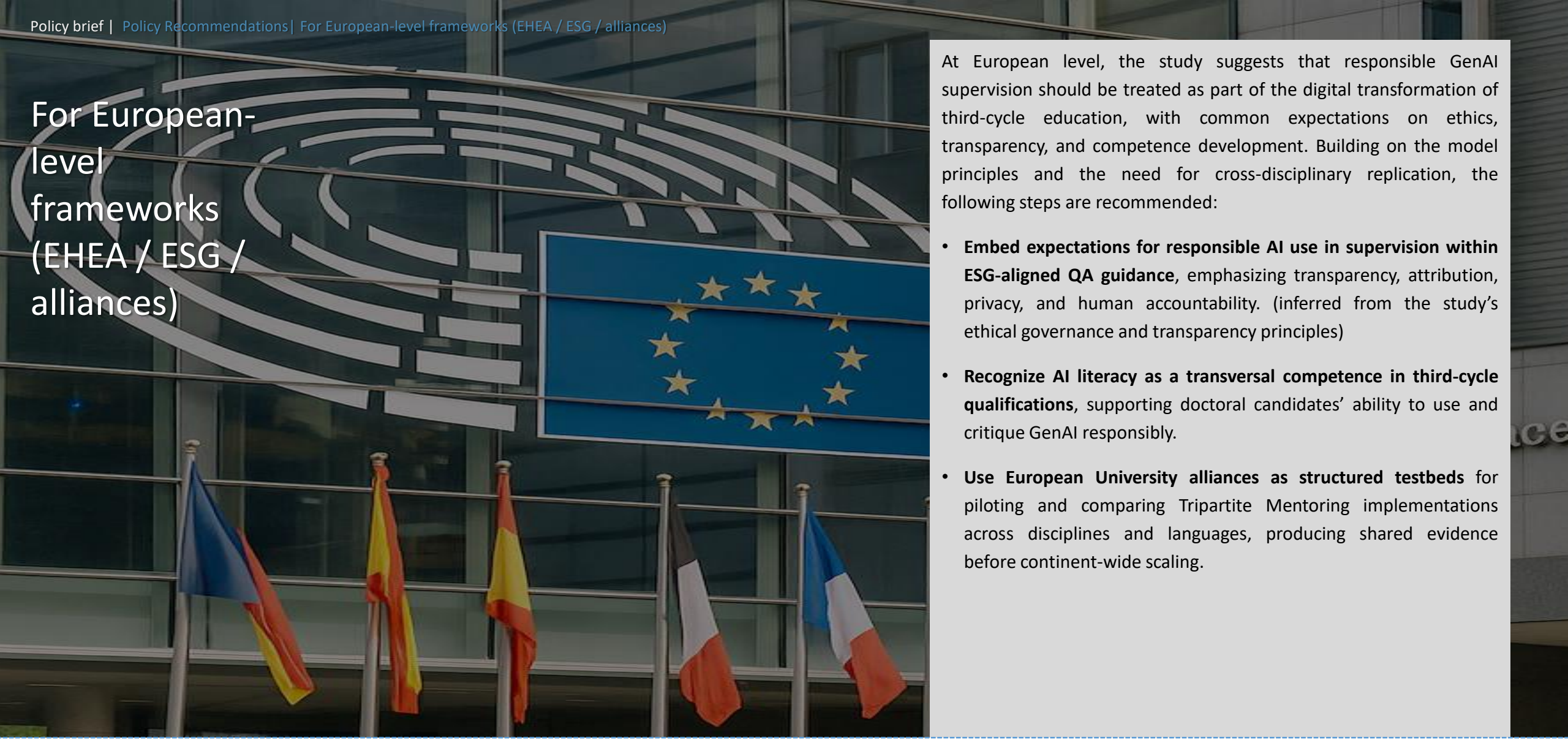
National authorities have a key role in ensuring that AI-assisted supervision develops within a trusted and comparable quality baseline across institutions. Based on the study's governance emphasis (ethics, literacy, accountability, and oversight), the following actions are recommended:

- **Update doctoral supervision standards to explicitly address GenAI-assisted practices,** recognizing AI as a bounded support tool while reaffirming the primacy of human scholarly judgment.
- **Support capacity-building programs for supervisors, doctoral schools, and QA personnel,** focused on AI literacy, prompt-based supervision methods, and critical evaluation of AI outputs.
- **Integrate AI-related criteria into national QA frameworks,** including requirements on disclosure, student agency, data safety, and monitoring of AI-supported guidance quality. (inferred from the study's ethical and accountability principles)

Universities should institutionalize AI literacy, integrity rules, and GDPR-safe practice before GenAI becomes supervision-by-default

The Tripartite Mentoring Model can be piloted within Universities as a scalable, QA-aligned blueprint for human-led AI-augmented supervision

For European-level frameworks (EHEA / ESG / alliances)



At European level, the study suggests that responsible GenAI supervision should be treated as part of the digital transformation of third-cycle education, with common expectations on ethics, transparency, and competence development. Building on the model principles and the need for cross-disciplinary replication, the following steps are recommended:

- **Embed expectations for responsible AI use in supervision within ESG-aligned QA guidance**, emphasizing transparency, attribution, privacy, and human accountability. (inferred from the study's ethical governance and transparency principles)
- **Recognize AI literacy as a transversal competence in third-cycle qualifications**, supporting doctoral candidates' ability to use and critique GenAI responsibly.
- **Use European University alliances as structured testbeds** for piloting and comparing Tripartite Mentoring implementations across disciplines and languages, producing shared evidence before continent-wide scaling.

Universities should institutionalize AI literacy, integrity rules, and GDPR-safe practice before GenAI becomes supervision-by-default

The Tripartite Mentoring Model can be piloted within Universities as a scalable, QA-aligned blueprint for human-led AI-augmented supervision

Risks and Study Limits

The study highlights that GenAI can strengthen doctoral guidance, but only under clear boundaries and critical human oversight. Several risks need to be actively managed if AI is embedded into supervision practices. First, **bias and uneven quality** may occur because AI systems inherit patterns from large training datasets and can surface partial or skewed perspectives; hence, outputs must be critically reviewed rather than accepted as authoritative.

Second, there is a **risk of over-reliance on AI** recommendations, especially by early-stage candidates who may interpret fluent text as correct or complete. This could weaken independent scholarly judgment and blur accountability for research decisions.

Third, **academic integrity and authorship boundaries** require explicit safeguards. AI should function as a cognitive support tool, not a co-author, and its contributions must remain transparent within supervision and in scholarly outputs.

Finally, **data privacy is a non-negotiable constraint**. Doctoral candidates should not input sensitive, identifiable, or unpublished data into public AI systems. Institutional rules need to ensure GDPR-aligned, safe use of GenAI in research contexts.

GenAI can enhance PhD supervision, but risks (bias, over-reliance, integrity, privacy) and the study's single-case scope make structured governance and further replication essential

Conclusion and Call to Action

Generative AI is already shaping doctoral work, and the question for higher education is no longer whether these tools will enter supervision, but **how** they can be integrated without eroding scholarly agency, rigor, or trust. Evidence shows that GenAI can provide doctoral-level guidance that experts deem appropriate, but **only under strong prompting structure and human critical mediation**. The large performance gap between structured/topic-specific prompts and generic keyword prompting demonstrates that AI usefulness is not automatic; it is a teachable, governable practice.

The Tripartite Mentoring Model offers a practical supervision blueprint for this new reality. By defining complementary roles for student, supervisor, and AI—and grounding their collaboration in explicit governance principles—it enables universities and alliances such as INVEST to benefit from GenAI's scalability while preserving the human, ethical, and intellectual core of doctoral training.

Call to action: Universities should adopt institutional AI-supervision guidelines, embed AI literacy in doctoral training, and pilot the Tripartite Mentoring Model within quality-assured frameworks. National QA agencies and European bodies should support these pilots, refine standards accordingly, and promote cross-disciplinary replication so that responsible AI-augmented supervision becomes a trusted and equitable component of doctoral education across the EHEA.

GenAI can strengthen doctoral supervision when governance, prompt literacy, and human oversight are built in from the start

Supporting Responsible AI Use in Doctoral Supervision

This policy brief is a University of Thessaly (UTH)–led contribution, developed within the INVEST European University Alliance and based on an INVEST-funded doctoral supervision case study. It draws from the peer-reviewed study "*Tripartite Mentoring in Doctoral Education: Evaluating Generative AI's Role in Supervision*", accepted for publication in the International Journal of Doctoral Studies (IJDS).

The brief supports EU priorities on academic excellence, digital innovation, and trustworthy AI integration in education and research. It also aligns with Bologna Process values and European quality-assurance frameworks.

This brief reflects UTH's findings and recommendations and does not constitute an official INVEST alliance position unless formally endorsed.

Research Theme: Digital Transformation in Doctoral Education

Affiliation: INVEST European University Alliance | University of Thessaly

Associated Journal: International Journal of Doctoral Studies (IJDS)

Contacts

Dr. Omiros Iatrellis
Assistant Professor, University of Thessaly, Greece
Email: iatrellis@uth.gr
https://ds.uth.gr/en/staff-en/faculty-en/iatrellis/

Prof. Nicholas Samaras
Professor, University of Thessaly, Greece
Email: nsamaras@uth.gr
https://ds.uth.gr/en/staff-en/faculty-en/iatrellis/

A European University Alliance
invest
The Competitive University for a Sustainable Society

Co-funded by the European Union