

AI Researchers

AI Researchers
Embedding Responsible AI in Research

AI ENGAGEMENT FRAMEWORK

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

CC BY SA



CONTENTS

1 **Executive summary** 3

2 **How we built the framework** 6

Introduction 7

Empirical foundation 10

Conceptual foundation of the framework 13

3 **AI engagement framework** 15

Framework overview 16

Intervention logic 20

Domain-specific application 24

Design principles 25

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

CC BY SA

IN THIS SECTION

Executive summary 4
Recommendations in a nutshell 5

01

EXECUTIVE SUMMARY



Executive summary

The AI Engagement Framework is a practical model for moving GenAI adoption in academic and research work from abstract debate to responsible, task-specific engagement. Its purpose is not to maximise AI use, but to help institutions, supervisors and researchers decide where AI creates visible value, where it requires control, and where it should not be used. The framework responds directly to five identified barriers: epistemic uncertainty, identity and role threat, knowledge gaps, institutional ambiguity, and task-specific resistance. Empirical findings show that researchers are more likely to engage with AI when they see concrete usefulness and compatibility with existing academic workflows; ease of use matters mainly because it helps users discover that value, while trust remains a constraint that must be built through verification and safeguards. Conceptually, the framework treats adoption as a cognitive, social and organisational process governed by the Human-AI Responsibility Boundary: AI may support process execution, but human judgement, ethics, evaluation, mentoring, authorship and accountability remain non-delegable.

Each barrier calls for a targeted response: Epistemic uncertainty calls for trust scaffolding through verification protocols and transparency mechanisms; identity and role threat calls for cognitive reframing by positioning AI as augmentation rather than replacement; knowledge gaps call for capability building through task-based training; institutional ambiguity calls for structural support through access, tools, and policies; and task-specific resistance calls for boundary setting regarding where AI should not be used, alongside workflow embedding through integration into everyday tasks. This logic turns AI engagement into a diagnostic practice rather than a generic technology campaign: identify the barrier in a concrete academic task, map the mechanism, select the intervention, apply it to a real use case, and reinforce it through policy, access, examples, verification and supervisory modelling.

Recommendations in a nutshell

The practical implication is clear: responsible engagement grows when AI is embedded where it improves process work, controlled where outputs affect academic quality, and excluded from domains where responsibility cannot be transferred.

RECOMMENDATION	IMPLICATION FOR ACTION
Focus on use cases	Start with concrete academic workflows where value is visible: literature review support, structured drafting, data preparation, content analysis, reporting and teaching-material preparation.
Embed AI into workflows	Design training and tools around existing research rhythms instead of presenting GenAI as an external disruption.
Target concrete barriers	Match interventions to specific barriers: trust scaffolding for uncertainty, cognitive reframing for identity threats, capability building for knowledge gaps, structural support for ambiguity, and boundary setting with workflow embedding for task-specific resistance.
Set responsibility boundaries	Explicitly protect human-only domains: ethics, evaluation, mentoring, judgement, intellectual ownership and sensitive relationships.

IN THIS SECTION

Introduction 7
Problem statement 9
Empirical foundation 10
Conceptual foundation of the framework 13

02

HOW WE BUILT THE FRAMEWORK



Introduction

Context: GenAI in academic work - between operational enthusiasm and intellectual scepticism

Generative AI has entered academic work as both a productivity tool and a source of normative tension. It can accelerate text drafting, summarise literature, process data, support reporting and generate teaching materials. But at the same time, it challenges established expectations around authorship, expertise, intellectual labour, reproducibility and responsibility.

The academic context makes these tensions especially significant. Research institutions are built on reliable knowledge, transparent methods and accountable judgement. While GenAI offer undeniable efficiency gains, their capacity to generate fluent but false outputs, compromise the confidentiality of unpublished research and blur the boundary between assistance and authorship demands a more rigorous evaluation framework.



What actually drives AI adoption?

Grounded in the extended **Technology Acceptance Model (TAM)**, evidence shows that AI adoption in academia is not driven by social pressure, but by rational utility. Researchers adopt GenAI based on two core predictors: perceived usefulness and perceived compatibility with their established workflows. Ease of use acts as a vital catalyst, lowering the cognitive cost of learning the system and quickly revealing its practical utility.

A tool, not a replacement

GenAI functions optimally as a high-quality administrative assistant, automating structured tasks like literature reviews and grant documentation preparation. This frees cognitive resources for deep critical synthesis. Crucially, AI democratises access to global science, helping non-native English speakers compete equally on the publication market. In conceptual phases, AI acts as an intellectual sparring partner, though final substantive verification and responsibility must remain unequivocally human.

The anatomy of distrust: barriers and non-negotiable limits

Trust in GenAI remains low due to strict academic standards. The boundary between delegable and non-delegable tasks is defined by the scientific ethos. While researchers accept delegating routine administrative burdens, literature reviews, and preliminary data processing, they strictly reject AI for tasks involving unpublished data or core manuscript concepts due to intellectual property risks. Furthermore, foundational analytical tasks are deemed non-delegable because automating them erodes the mentor-apprentice relationship crucial for junior scholars. Finally, a non-negotiable red line is drawn around relational and ethical domains: tasks requiring moral judgement, conflict resolution, student assessment, and individual mentoring cannot be outsourced to algorithms without degrading the process.

The paradox of responsibility & strategic implications

A compelling paradox emerges among institutional leaders: while they enthusiastically adopt AI to optimise managerial and bureaucratic burdens, they simultaneously set the hardest limits against delegating abstract, ethical tasks. Strategically, institutions must shift from generic training to task-oriented AI integration. To prevent severe access inequality, universities must provide secure, auditable access to advanced tools. Finally, coherent regulatory clarity is urgently needed to establish transparent ethical frameworks for responsible deployment.

METHODOLOGICAL NOTE

This AI Engagement Framework is based on a dual-phase research study conducted in early 2026, the results of which were presented in the **AI Adoption Barriers Report**. The findings incorporate quantitative data from 136 researchers in Poland, Spain, and Portugal (using Structural Equation Modelling based on an extended Technology Acceptance Model) and qualitative insights from four Focus Group Interviews with academic leaders and supervisors in Poland, Spain, and Portugal.

Problem statement

The core problem is not simply low adoption of a new technology. The evidence suggests a more specific pattern: researchers can see the potential of GenAI, but they remain cautious, selective and boundary-conscious. Adoption is filtered through perceptions of usefulness, compatibility, trust, institutional legitimacy and professional identity.

Barriers therefore operate at several levels. Some are **epistemic**, such as hallucinations, instability and non-reproducibility. Some are **institutional**, such as unclear policies and informal prohibitions. Some are **identity-related**, including fear of devalued expertise or AI being interpreted as cheating. Others are **task-specific**, emerging when AI enters domains of judgement, evaluation, ethics, mentoring, privacy or intellectual property.

Purpose of the framework

The framework uses these barriers to build the operational logic upon them. It supports institutions and supervisors in designing AI engagement that is evidence-based, task-specific and aligned with academic values. It also positions AI as research enhancement rather than replacement: a tool that can extend capacity where work is processual, but must remain subordinate to human responsibility where meaning, judgement and accountability are at stake.



It also positions AI as research enhancement rather than replacement.

Purpose of the framework

Empirical foundation

Qualitative insights

Focus group findings show that academic concerns about GenAI are multidimensional. Participants did not reject AI as such; rather, they differentiated between acceptable forms of assistance and unacceptable forms of substitution. This distinction is the empirical basis for the framework.

Epistemic uncertainty

FRAMEWORK IMPLICATION

Adoption requires verification routines and explicit human review.

Hallucinations, polished falsehoods, unstable outputs and limited reproducibility undermine scientific confidence.

"My experiences have been rather negative. I tested a tool for generating scientific texts that was supposed to use reliable databases. In practice, it turned out to generate a large number of hallucinations. The biggest problem, however, was that the interpretations were often disconnected from the original content of the articles. After verifying the sources, it turned out that the conclusions were inadequate."

Identity and role threat

FRAMEWORK IMPLICATION

Position AI as augmentation and preserve scholarly authorship.

AI can be perceived as devaluing expertise, weakening apprenticeship and blurring legitimate intellectual effort.

"I look at this from a broader perspective, not just through the lens of research quality, but more philosophically. The key question I ask myself is: what skills do I want to preserve? If we delegate too many tasks to AI, these skills naturally begin to fade—just as the ability to read maps did after the advent of GPS navigation. For me, it's important to consciously decide what I want to keep doing on my own. For example, I don't want to hand over the writing process to AI—not just because of quality, but because I want to preserve my own style and creative freedom."

Knowledge gaps

FRAMEWORK IMPLICATION

Provide task-based capability building instead of generic awareness sessions.

Researchers are uncertain about available tools, practical use cases and safeguards.

"For me, the key issue is access to tools. Right now, many of us purchase them on our own, but the university should provide access and, at the same time, teach us how to use them. If we want to teach students, we ourselves must have the appropriate tools and competencies. This should not rely solely on individual initiatives."

Institutional ambiguity

FRAMEWORK IMPLICATION

Define permitted, conditional and prohibited uses.

Unclear rules and informal prohibitions create anxiety, especially for junior researchers.

"let's say a professor who's 20 years older doesn't use AI because he's put off by it, considers it unethical, and forbids me from using it. So – perhaps some general guidelines like that would be useful."

Task-specific resistance

FRAMEWORK IMPLICATION

Use a responsibility boundary to protect human-only domains.

Ethics, mentoring, evaluation, conflict resolution and IP/privacy are treated as non-delegable.

"The key is making a conscious decision about which competencies we develop ourselves and which we support with tools. I also see significant risks in the areas of ethics and the quality of knowledge. There is a risk of "dumbing down" information—the dissemination of content that is not well-verified but functions as if it were true. I'm also worried that we might lose our critical thinking and sense of responsibility for content."

The quotations were derived from Focus Group Interviews with academic leaders and supervisors in Poland, Spain, and Portugal.

Quantitative model: extended Technology Acceptance Model (TAM)

Perceived usefulness and perceived compatibility are the two direct predictors of behavioural intention. This means researchers are most likely to engage with GenAI when **they believe it solves real professional problems** and **fits the way they already work**.

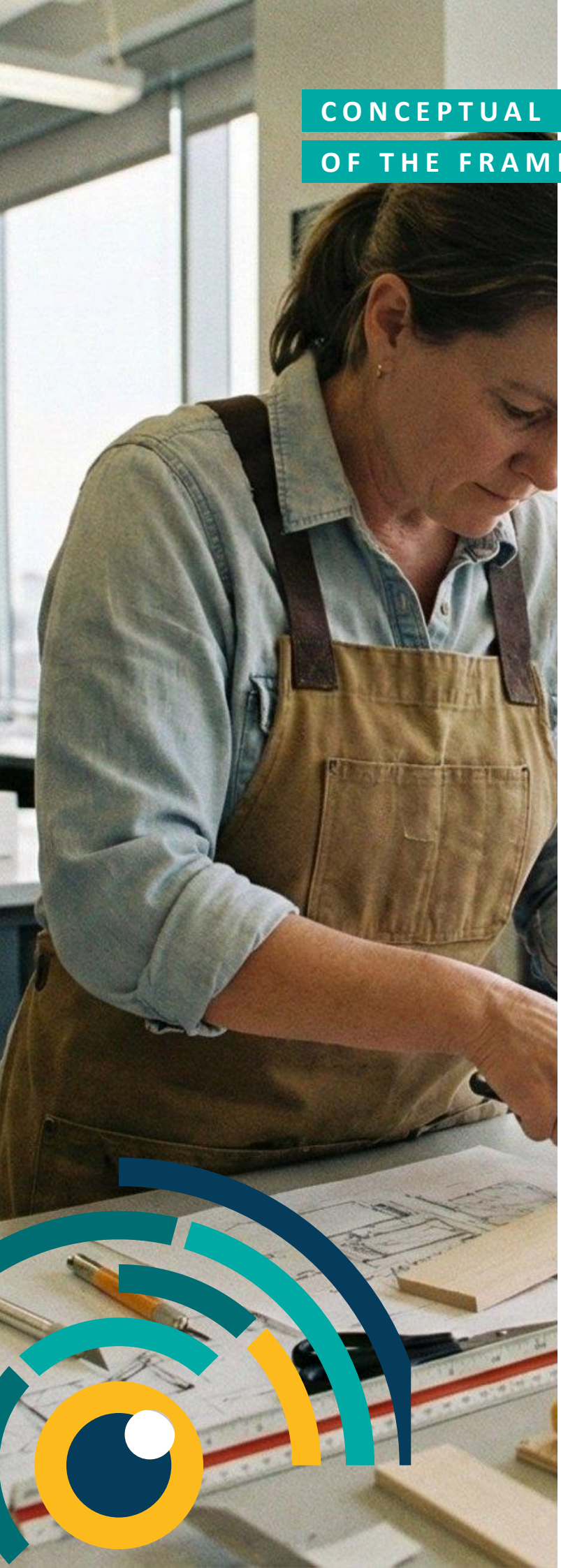
Perceived ease of use strongly shapes perceived usefulness. In practical terms, intuitive tools matter because they reduce the cognitive cost of experimentation and help researchers reach the point where they can evaluate value for themselves.

Trust and social influence require a careful interpretation. The average level of trust is descriptively low, indicating a gap between moderate willingness to use GenAI and confidence in its reliability and safety. Social influence is also relatively weak, which suggests that academic engagement with GenAI is not primarily driven by peer pressure or institutional mandates, but rather by perceived personal and practical value — the stronger the belief in concrete use cases, the higher the perceived value of GenAI.

Matrix-based insights

The matrix findings provide the clearest empirical expression of the Human-AI Responsibility Boundary. Across domains, high acceptance appears where tasks are repetitive, data-intensive, formalised or process-oriented. Low acceptance appears where tasks require ethical judgement, relational sensitivity, individual assessment, accountability, original thought or legal responsibility.

HIGH ACCEPTANCE	CONDITIONAL ACCEPTANCE	RESTRICTED DOMAINS
Data analysis	Academic writing	Ethics
Summarisation	Literature synthesis	Evaluation
Formatting	Feedback support	Mentoring
Project reports	Research-method selection and interpretation support when expert verification remains central	Conflict resolution
Grant drafting		Strategic judgement
Research documentation		Privacy
Teaching-material preparation		Intellectual property
		Final accountability



CONCEPTUAL FOUNDATION OF THE FRAMEWORK

Adoption as a process

The AI Engagement Framework was developed in response to the barriers identified in the AI Adoption Barriers Report. Adoption is not a binary state: a tool is either used or not used. Adoption is the process through which researchers evaluate, test, negotiate and embed AI in their work while preserving professional judgement.

Adoption was recognised as having cognitive, social and organisational dimensions.

Cognitively, researchers need to **recognise value and understand limits**.

Socially, they need **legitimate norms** for discussing and modelling AI use.

Organisationally, they need **access, policies, secure infrastructure and training** that fit their real workflows.

Human-AI Responsibility Boundary

The Human-AI Responsibility Boundary is the signature element of the framework. It defines which forms of work can be supported or partially delegated to AI, and which forms must remain under human authority. The boundary is not anti-technology; it is a design principle for responsible engagement (Figure 1).

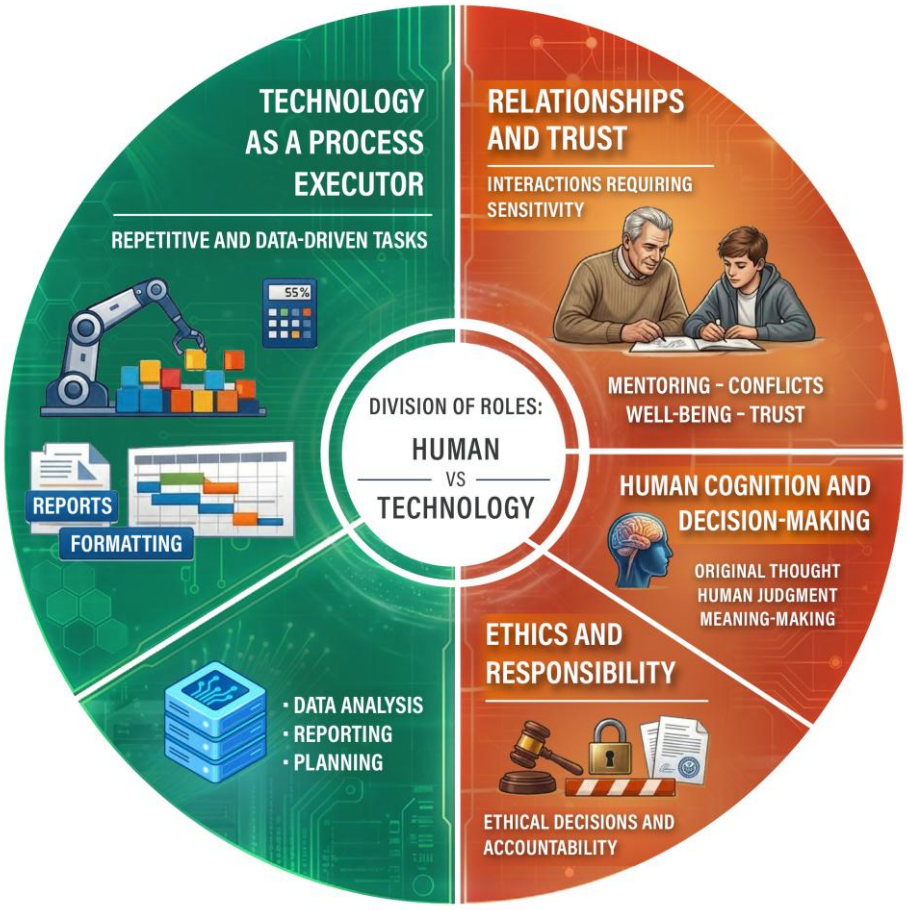


Figure 1. Visual summary of the Human-AI division of roles: process execution versus human responsibility.

On the delegable side are tasks where AI acts as a process executor: formatting, reporting, data analysis, summarisation, planning support and administrative drafting. On the human-only side are relationships, trust, mentoring, ethics, original meaning-making, judgement, evaluation and accountability. The framework uses this division as a guardrail for intervention design.

IN THIS SECTION

Framework overview 16
Framework diagram 18
Intervention logic 20
Domain-specific application 24
Design principles 25

03

AI ENGAGEMENT FRAMEWORK



Framework overview

The framework integrates findings from both the qualitative FGI study and the quantitative TAM-based analysis. The barriers were identified through focus group interviews and reflect the contextual challenges of AI adoption in researchers' everyday work. The core mechanisms are grounded in the quantitative model, where Perceived Usefulness and Compatibility emerged as the primary drivers of AI adoption, supported by Ease of Use. The connections between barriers, mechanisms, and interventions represent an interpretative synthesis, combining empirical findings with theory-informed practical implications for supporting responsible AI adoption in research environments.

The framework is organised as a practical transformation chain across three layers. Layer 1 (Barriers) captures the empirically observed barriers. Layer 2 (Mechanisms) explains why these barriers affect engagement through the TAM core. Layer 3 (Interventions) translates the mechanisms into concrete institutional and supervisory actions (Figure 2).



Layer 1: Barriers

The barriers layer prevents implementation work from starting with technology features. Instead, it begins with the reasons researchers hesitate, refuse or conditionally accept AI. This is important because different barriers require different responses: hallucinations require verification, identity threat requires reframing, and institutional ambiguity requires rules.

Layer 2: Mechanisms

At the centre of the framework are perceived usefulness and compatibility. Perceived usefulness increases when researchers recognise concrete gains in productivity and research quality, addressing the question of whether GenAI helps them achieve academic goals more effectively. Compatibility, in turn, reflects whether these tools can be integrated into existing academic workflows and values without undermining professional standards in research, supervision, teaching, writing, or project management.

Ease of use is treated as an enabling condition rather than a final objective. A simple and accessible interface matters primarily when it helps researchers identify credible and meaningful use cases. Trust, meanwhile, is approached not as a message to be promoted, but as a contextual constraint that must be carefully scaffolded through transparency, verification mechanisms, and bounded delegation.

Layer 3: Interventions

The intervention layer defines six action types: trust scaffolding, cognitive reframing, capability building, structural support, boundary setting and workflow embedding. Together, these actions turn the framework from an analytical model into an implementation logic.

Trust scaffolding

Cognitive reframing

Capability building

Structural support

Boundary setting

Workflow embedding

Framework diagram

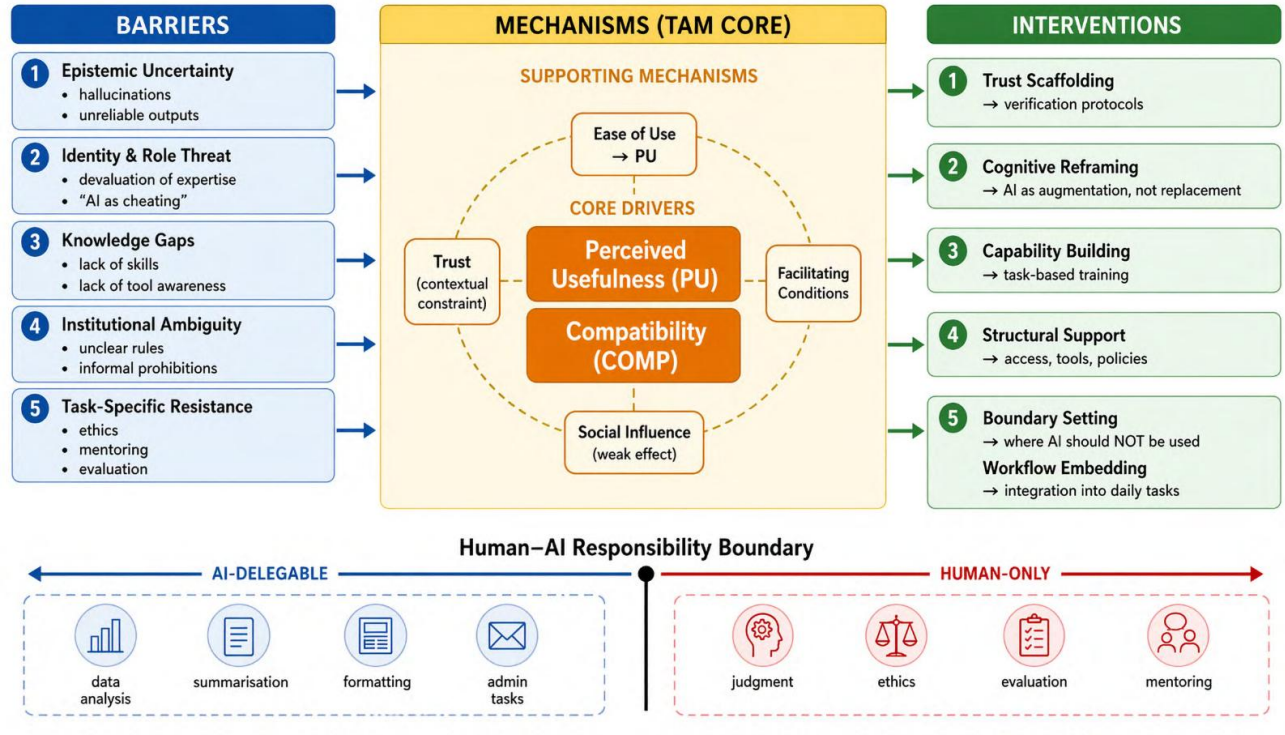


Figure 2. AI Engagement Framework: from barriers, through Technology Acceptance Model (TAM) mechanisms, to interventions and the Human-AI Responsibility Boundary.

The diagram should be read from left to right and then through the boundary at the bottom. The left column summarises the main insight categories: epistemic uncertainty, identity and role threat, knowledge gaps, institutional ambiguity and task-specific resistance. These are not isolated obstacles; they are recurring sources of hesitation that shape how researchers interpret GenAI.

The central mechanism translates those barriers into the TAM core. Perceived usefulness and compatibility are the main drivers. Ease of use supports perceived usefulness. Trust, facilitating conditions and social influence are placed around the core because they shape the context in which usefulness and compatibility can become credible.

The right column turns the model into action. Trust scaffolding responds to epistemic uncertainty through verification protocols and transparency mechanisms; cognitive reframing responds to identity and role threat by positioning AI as augmentation rather than replacement; capability building responds to knowledge gaps through task-based training; structural support responds to institutional ambiguity through access, tools, and policies; and boundary setting together with workflow embedding responds to task-specific resistance by defining where AI should not be used and integrating it into everyday tasks.

The responsibility boundary at the bottom is the normative anchor of the framework. It prevents the intervention logic from becoming a simple adoption funnel. The goal is not maximum delegation, but appropriate delegation: AI-supported process work where value is high, and human authority where responsibility, judgement and relationships are central.



Intervention logic

From barrier to action

The proposed framework works as a diagnostic sequence. Institutions and supervisors should first identify the barrier, then map it to the mechanism it affects, select the relevant intervention, apply it to a concrete use case and reinforce the new practice through workflow support.

1 **Identify the dominant barrier in a specific academic task.**

2 **Map the barrier to the mechanism it weakens: usefulness, compatibility, ease of use, trust.**

3 **Select the intervention type that addresses that mechanism.**

4 **Apply the intervention to a concrete use case rather than to AI in general.**

5 **Reinforce the practice through policy, access, examples, verification and supervisory modelling.**

Types of interventions

Trust scaffolding

WHEN TO USE IT

Outputs are seen as unreliable, unverifiable or risky.

INSTITUTIONAL LEVEL IMPLEMENTATION

Introduce verification protocols, source-checking procedures, mandatory human review requirements, and guidance on documenting AI-assisted work.

EXAMPLE ACADEMIC USE CASE

Researcher uses AI to summarise 50 articles, but all references and interpretations are checked against the original sources before inclusion in a literature review.

EXPECTED EFFECT

Builds confidence through verification, traceability and review protocols.

Cognitive reframing

WHEN TO USE IT

AI is perceived as cheating, replacement.

INSTITUTIONAL LEVEL IMPLEMENTATION

Develop institutional communication that frames AI as research support rather than academic misconduct; provide examples of responsible AI-assisted work.

EXAMPLE ACADEMIC USE CASE

Supervisor demonstrates using AI to generate alternative structures for a grant proposal while retaining full authorship and decision-making.

EXPECTED EFFECT

Legitimises AI as augmentation while preserving academic agency.

WHEN TO USE IT

AI is perceived as devaluation of expertise.

INSTITUTIONAL LEVEL IMPLEMENTATION

Emphasise preservation of scholarly authorship, critical thinking and expert judgement in AI policies and training materials.

EXAMPLE ACADEMIC USE CASE

Researcher uses AI to generate brainstorming prompts but develops the conceptual framework and theoretical contribution independently.

EXPECTED EFFECT

Reinforces professional identity while positioning AI as a complementary tool.

Types of interventions

Capability building

WHEN TO USE IT

Researchers lack practical knowledge of tools and use cases.

INSTITUTIONAL LEVEL IMPLEMENTATION

Deliver task-based training, discipline-specific workshops, practical use-case libraries and peer-learning initiatives.

EXAMPLE ACADEMIC USE CASE

Workshop on AI-assisted verbatim qualitative coding, literature screening, and academic writing support for doctoral researchers.

EXPECTED EFFECT

Raises ease of use and helps users discover usefulness.

Structural support

WHEN TO USE IT

Access or subscriptions are inconsistent.

INSTITUTIONAL LEVEL IMPLEMENTATION

Provide institution-wide access to approved AI tools and equitable subscription models.

EXAMPLE ACADEMIC USE CASE

University provides institution-wide access to approved GenAI tools for all researchers and doctoral students.

EXPECTED EFFECT

Makes responsible engagement equitable and institutionally credible.

WHEN TO USE IT

Security and policies are inconsistent.

INSTITUTIONAL LEVEL IMPLEMENTATION

Establish clear AI governance, approved-tool lists, data-protection standards, privacy guidance and IP policies.

EXAMPLE ACADEMIC USE CASE

Research team uses an institutionally approved AI platform for analysing interview transcripts under established data-protection guidelines.

EXPECTED EFFECT

Reduces uncertainty and increases trust in institutional support.

Types of interventions

Boundary setting

WHEN TO USE IT

Users are uncertain where AI is appropriate or prohibited.

INSTITUTIONAL LEVEL IMPLEMENTATION

Publish a "Permitted–Conditional–Prohibited" use framework and define human-only domains requiring human responsibility.

EXAMPLE ACADEMIC USE CASE

Faculty guidance specifies that AI may support feedback drafting but cannot determine student grades or final assessment decisions.

EXPECTED EFFECT

Reduces ambiguity and protects human-only domains.

Workflow embedding

WHEN TO USE IT

AI is seen as external or extra work.

INSTITUTIONAL LEVEL IMPLEMENTATION

Integrate AI tools into existing research, teaching and project-management workflows; provide workflow-specific templates and examples.

EXAMPLE ACADEMIC USE CASE

AI-assisted literature screening is integrated into the standard systematic review workflow and research-methods training.

EXPECTED EFFECT

Improves compatibility with existing academic routines.

WHEN TO USE IT

AI is seen as disruptive work.

INSTITUTIONAL LEVEL IMPLEMENTATION

Introduce AI gradually through pilot projects, local champions and demonstrated use cases that show value within existing practices.

EXAMPLE ACADEMIC USE CASE

Department pilots AI-supported project reporting and grant-administration processes before scaling adoption across research groups.

EXPECTED EFFECT

Reduces perceived disruption and increases acceptance through visible usefulness.

Domain-specific application

High-acceptance domains

High-acceptance domains are the natural starting point for implementation. They include data analysis, summarisation, formatting, administrative tasks, reporting, grant drafting, research documentation and selected teaching-material preparation. These tasks are attractive because value is visible and responsibility can remain clearly with the researcher.

Use AI to accelerate repetitive and formalised work, not to remove expert oversight.

Frame productivity gains as time returned to interpretation, supervision and scientific judgement.

Provide examples within real academic workflows, such as literature screening, report drafting or slide preparation.

Conditional domains

Conditional domains require a more careful design. Academic writing, teaching support, feedback preparation, literature synthesis, method selection and interpretation support can benefit from GenAI, but only under explicit human verification. In these domains, AI should support structure, alternatives and preliminary synthesis; it should not become the source of final claims.

Require disclosure and verification norms where authorship or assessment could be affected.

Use AI outputs as drafts, prompts or structured inputs rather than final academic products.

Make the point of human responsibility visible in guidelines and training materials.

Restricted domains

Restricted domains are those where the empirical evidence shows clear resistance and where professional responsibility cannot be transferred to a system. These include ethics, assessment of student work, mentoring, conflict resolution, strategic judgement, privacy, intellectual property, research integrity and final accountability for findings.

Do not delegate moral judgement, relational responsibility or final evaluation to AI.

Do not process sensitive, unpublished or proprietary research material through tools without approved security guarantees.

Use AI, if at all, only for bounded administrative support around these domains, never as the decision-maker.

Design principles

The principles below constitute a coherent scaffolding for building researcher engagement with AI. Taken together, they provide a solid foundation for the adoption of AI as a tool for augmenting research capabilities while remaining aligned with academic values, human judgement, and research ethics. They translate the empirical findings into practical rules for the development of future tools, training, policies, and toolkit materials.

Use-case first	Start from the academic task, not from the technology or a generic prompt catalogue.
Augmentation, not replacement	Frame AI as a support for research capacity while keeping authorship and judgement human.
Respect non-delegable domains	Protect ethics, relationships, evaluation, mentoring and accountability from automation.
Make usefulness visible	Show concrete benefits in familiar workflows and allow researchers to test value directly.
Embed, do not disrupt	Integrate AI into existing routines, tools and institutional practices.
Trust through control	Build confidence with verification, transparency and data safeguards.
Bottom-up adoption	Prioritise locally meaningful examples over top-down pressure.

AI Researchers

Embedding Responsible AI in Research

Use, share and adapt

This framework is published under a Creative Commons Attribution-ShareAlike 4.0 licence (CC BY-SA 4.0). You may share and adapt it for any purpose, provided you credit the source and share your adaptations under the same licence.

airesearchers.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

CC BY SA