EPiC Series in Computing

Volume 109, 2026, Pages 231–240

Proceedings of EUNIS 2026 Annual Congress

Embracing AI-Driven Change in Education: A Student–Instructor Centered Institutional Framework

Lisa Cosaro[1*] and Nicola Russo[1†]

[1] Bocconi University, Milan, Italy

lisa.cosaro@unibocconi.it, nicola.russo2@unibocconi.it

Abstract

The rapid evolution of generative Artificial Intelligence is reshaping the higher education sector, creating new pedagogical opportunities for teaching and learning while simultaneously raising significant governance, and regulatory challenges. While commercial AI tools are widely accessible, universities require institutionally governed solutions that ensure data sovereignty, pedagogical control, and compliance.

This paper introduces a student–instructor centered framework to AI adoption in higher education, grounded in the institutional design and deployment of a fully in-house developed AI platform. The initiative was developed through a structured co-design process involving faculty members in pilot implementations, structured feedback collection, and iterative prioritization of system enhancements.

The resulting platform integrates generative AI into teaching within a secure and controlled ecosystem: all data are internally managed, hosted within the European Union, and excluded from external training or profiling purposes. Through Retrieval-Augmented Generation (RAG), instructors can configure course-specific AI agents, define behavioral parameters, and constrain the knowledge perimeter to reduce hallucinations and ensure contextual alignment.

The model further emphasizes instructor oversight and institutional governance. Faculty retain visibility over student–AI interactions, access usage analytics and conversation exports, and collect student feedback. At the institutional level, embedded guardrails prevent high-risk practices such as automated grading, aligning the system with the European AI Act's risk-based approach.

By combining pedagogical co-design, institutional control, technological robustness, and regulatory alignment, this work proposes a replicable framework for secure, governed, and student–instructor centered AI integration in higher education.

[*] Author of this paper and co-author of EDA

[†] Reviewer of this paper and co-author of EDA & LUIGI

L. Desnos, C. Diaz, J. Mincer-Daszkiewicz, L. Merakos, R. Vogl, S. McLellan and U. Lucke (eds.), EUNIS 2026 (EPiC Series in Computing, vol. 109), pp. 231–240

1 Introduction

The adoption of commercial AI tools outside formal institutional oversight generates a complex and multi-layered landscape of governance, pedagogical, and regulatory challenges.

First, when students upload instructors' copyrighted materials or sensitive personal data to third-party AI platforms, it is often unclear whether they are retained into model training or to what extent they may be accessed or reused by service providers, thereby posing significant risks to **intellectual property rights** and **data protection** (Jin et al., 2025).

Second, the use of commercial AI services enables students to access massive amounts of information from across the open web, as well as potentially outdated training data from large language models. When such information is not grounded in verified knowledge and appropriate context, these systems are highly likely to generate responses containing errors or hallucinations, and to produce outputs that do not align with course objectives. Such limitations pose significant **pedagogical concerns** (Qian, 2025).

Finally, in the European context, universities must comply with the requirements of the Artificial Intelligence Act, which adopts a risk-based approach to AI governance. According to Annex III, AI systems used in the following educational contexts, among others, are considered high-risk:

(b) AI systems intended to be used to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels (Regulation (EU), 2024)*;*

(c) AI systems intended to be used for the purpose of assessing the appropriate level of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions at all levels (Regulation (EU), 2024)*;*

In this complex landscape, institutions are compelled to balance the pedagogical potential of generative AI with institutional control and regulatory responsibility, ensuring that innovation does not compromise academic integrity or legal compliance.

2 A Co-Design, User-Centered Development Process

The development of the institutional AI platform followed a gradual approach, guided by end user requirements, institutional constraints, and iterative testing and refinement.

The initiative began with the creation of a controlled "playground" environment for early prototyping, named LUIGI (Lab for University Innovation with GenAI Ideas), conceived as a secure experimental space to explore generative AI conversational capabilities. The project emerged from internal research and development activities within the University. From the beginning, faculty members and staff were involved in testing initial functionalities within this low-risk environment.

Figure 1: LUIGI logo

The playground then evolved into a series of spin-off, course-specific instances, deployed on demand to address concrete instructional needs emerging from pioneering instructors and involving students for the first time. These spin-offs proved the potential of the technology and paved the way for broader institutional expansion, namely the integration into the university's Learning Management System (LMS). The integration of LUIGI Core into the LMS resulted in the development of the *EDTech AI Assistant (EDA)* also known as the AI Course Assistant, an AI-powered solution that enables instructors to design and deliver course-contextualized assistants to support students in several learning scenarios.

Before full-scale release, a semester-long piloting phase was conducted to validate usability, pedagogical alignment, and risk mitigation mechanisms at scale. During this phase, instructors were closely supported throughout their use of the platform: structured feedback was systematically collected to identify necessary improvements, and continuous interaction with the pedagogical team ensured methodological alignment. Platform usage was actively monitored to observe student behavior and interaction patterns, enabling data-informed refinements.

At the end of the semester, all participating students were invited to complete structured surveys to provide comprehensive feedback on their experience.

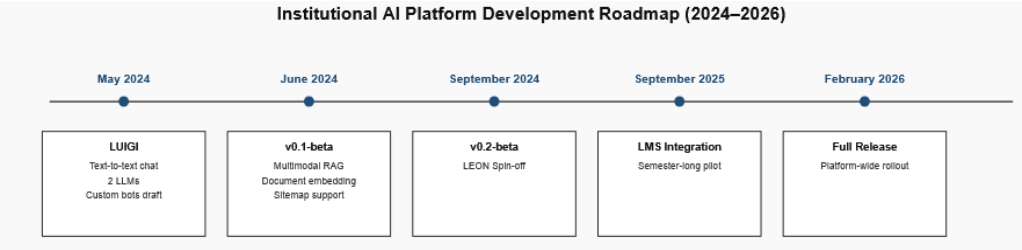

Figure 2: Timeline

Throughout the entire evolution process, extensive benchmarking activities were conducted, analyzing existing tools available on the market to assess whether a suitable solution already existed. The scouting revealed that, at the initial stage (early months of 2025), no tool could be seamlessly integrated into the LMS via LTI Standard. Moreover, many of the available solutions either delegated full control to platform administrators rather than instructors or offered only minimal levels of customization. This lack of flexibility and instructor-centered control ultimately reinforced the need to develop a dedicated solution tailored to our specific needs.

3 System Architecture and Technical Design Choices

An **in-house solution** allows to address one of the main challenges outlined in the introduction: ensuring that institutional data remain within the university's controlled perimeter.

The adopted solution is built on AWS managed services, leveraging General Purpose models, also known as Large Language Models (LLMs), available through Amazon Bedrock under a pay-per-token consumption model. The model selection also includes latest OpenAI GPT models, accessed via official OpenAI APIs using SDK. These APIs are provided through an EDU License following an agreement between Bocconi and OpenAI.

The LLM layer interacts with a Retrieval-Augmented Generation (RAG) architecture, where course materials are converted into embeddings and stored in a vector database to enable semantic retrieval at query time. This configuration allows the system to generate domain-specific

conversational bots whose responses are grounded in instructor-curated materials and constrained within the boundaries of each academic course.

Large Language Models (LLMs) are known to generate hallucinations, i.e., fluent but factually incorrect or unsupported statements, due to their reliance on parametric memory and probabilistic text generation (Ji et al., 2023).

RAG has been shown to improve factual accuracy and reduce unsupported claims in knowledge-intensive tasks (Lewis et al., 2020).

By constraining outputs within instructor-curated materials, RAG reduces epistemic drift and limits domain-inconsistent responses. While it does not eliminate hallucinations entirely, retrieval grounding significantly lowers their likelihood and enhances traceability of generated content (Ji et al., 2023) (Lewis et al., 2020).

The entire architecture — including retrieval mechanisms, embeddings, storage, and logging — is designed and governed internally and ensures scalability to the entire Institution (~ 45k users). All services are hosted within the European Union, ensuring compliance with European data protection standards.

Moreover, the pay-per-token consumption model allows for predictable and contained operational costs, making the solution economically sustainable on a scale.

A dedicated middleware layer manages course-context propagation and identity federation between the LMS and the core infrastructure. The integration adheres to the **LTI 1.3 standard**, leveraging its OAuth 2.0 and OpenID Connect-based security framework. Through the LTI launch flow, the middleware extracts the course context and participant roles, enabling dynamic configuration of the assistant instances associated with that specific course. Course materials are seamlessly synchronized with the assistant's Knowledge Base.

Authentication and authorization are also handled within this layer. LMS users are mapped to Amazon Cognito, where their identities are federated and associated with appropriate authorization policies. This mapping process enables users to securely access AWS services according to their institutional role and permissions.

The architecture diagram illustrates the core logical structure of the system, composed of 6 main layers: LTI middleware, frontend, backend API, LLM and RAG layer, and analytics/monitoring.

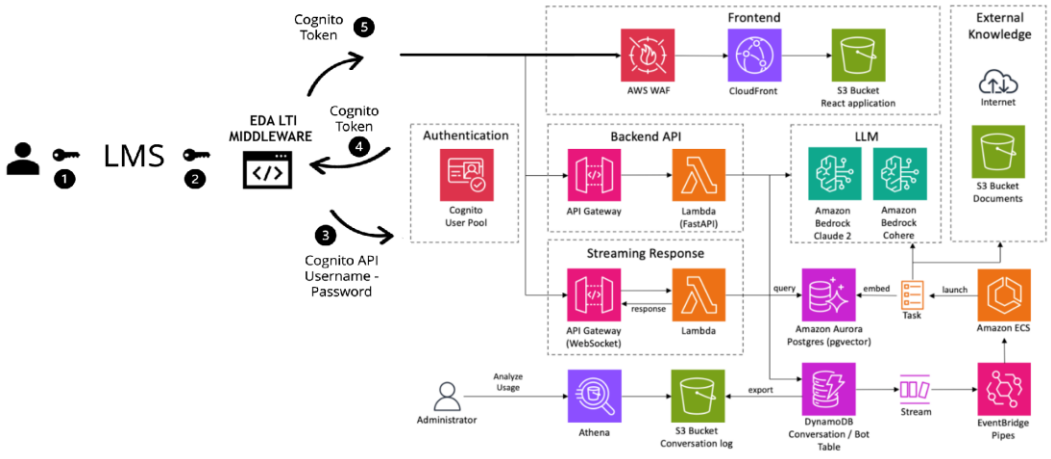

Figure 3: Technical Solution

4 Control, Transparency and Oversight for Instructors

The distinctive element of the proposed framework, compared to commercial AI tools, lies in its instructor-centered configurability and observability. The system does not treat AI as an opaque black box, but rather as a configurable technology operating under faculty control and institutional governance.

4.1 Configurable AI Assistants and Advanced RAG Configuration

Each assistant is defined through a detailed configuration process that combines behavioral design and knowledge base synchronization.

1. Instructors configure each assistant through a structured meta-prompt that defines its pedagogical role, objectives, tone, and communication style. The assistant may thus be intentionally designed as a Socratic tutor, a reflective facilitator, a study companion, or a logistical support agent. This behavioral alignment is explicit: the assistant does not autonomously determine its educational function, but operates within boundaries defined by the instructor. In practice, the most common use cases cluster around three primary categories. First, the assistant serves as a **study coach**, guiding reflection and clarifying complex concepts. Second, it provides **logistical assistance** by addressing questions related to deadlines, assignments, and syllabus content. Third, it supports **formative self-assessment** by generating practice questions and quizzes to self-evaluate understanding.
   Beyond these core applications, more advanced pedagogical scenarios can be developed in collaboration with the institutional methodological and pedagogical support team, further enhancing the quality and depth of the learning experience.

2. The Knowledge Base is entirely instructor controlled. Instructors select course materials from the LMS, choosing among various supported file formats, upload additional documents, and, where appropriate, integrate external web resources. This turns the assistant into a course-specific expert grounded in curated academic content rather than an open-domain conversational system.

3. The RAG configuration further strengthens this instructional control. Instructors can choose how documents are processed during the embedding phase. In Single Mode, a document is treated as a unified semantic unit, preserving its overall structure and coherence. In Paginated Mode, longer documents are segmented into structured sections, improving retrieval precision and allowing the assistant to focus on localized content. This choice directly influences how knowledge is accessed and contextualized in responses.
   When documents contain images, two processing strategies are available. Text Mode extracts visible text embedded within images, enabling semantic search over that content. Anchor Mode preserves visual structures such as diagrams or forms, allowing them to be referenced in responses. Advanced embedding configurations add another layer of customization, including chunk size and overlap, by which advanced users influence the granularity of semantic retrieval. Smaller chunks increase specificity, while larger ones preserve broader context. Overlap ensures continuity between adjacent segments, mitigating fragmentation.

4. The system also allows instructors to define custom conversation starters. These structured prompts appear as suggested entry points for students and guide interaction

toward reflection, application, or conceptual exploration. In this way, the assistant is not merely reactive but pedagogically scaffolded.

5. Model flexibility further reinforces instructor control. Faculty can experiment with different foundation models (latest Anthropic Claude Sonnet and OpenAI GPT models) and select a default model for student interaction. This enables comparative testing and ensures consistency in student experience once a default configuration is established.

4.2 Evaluation Lab and Testing

Before publishing assistants thus making them available to students, instructors are encouraged to deeply test them. The integrated **Evaluation Lab** feature introduces a structured approach to RAG quality assurance. This feature enables the generation of automated synthetic question–answer sets derived from the Knowledge Base, as well as the upload of manually curated datasets containing questions and expected answers. The system evaluates the assistant's responses, assigns performance scores, and allows inspection of the retrieved documents used during generation.

This allows for an iterative refinement cycle in which instructors configure the assistant, test its outputs, analyze discrepancies, adjust instructions or materials, and retest performance. AI deployment is thus reframed as a measurable and auditable process aligned with academic standards of validation. Rather than passively accepting generative outputs, faculty actively verify and calibrate the system's behavior.

4.3 Full Visibility and Monitorings

A key feature of the framework is advanced monitoring. The AI Course Assistant provides full visibility into student–AI interactions. Instructors can monitor conversations and export interaction logs. Filtering by date or feedback enables targeted analysis, and student identities may be disclosed when necessary for pedagogical support, within the boundaries established by institutional Terms and Conditions governing data protection and privacy.

Faculty can identify recurring misunderstandings, detect misuse, and assess how students engage with course content.

Instructors can also access the Analytics section providing a comprehensive overview of assistant usage through key metrics and dynamic charts.

4.4 Faculty Training and AI Literacy

Technical configurability alone does not guarantee meaningful instructor control; such control ultimately depends on adequate AI literacy.

Understanding the probabilistic nature of generative models, the limitations of LLMs, and the functioning of retrieval mechanisms is essential for informed use.

For this reason, the institution developed training initiatives, including training cycles and asynchronous learning materials, aimed at fostering critical and pedagogical awareness. In addition, one-to-one support channels were established to guide instructors throughout experimentation and the broader transition toward AI-enhanced teaching practices. Only when faculty understand both the tool and its underlying logic can they truly exercise control, leveraging AI strategically.

4.5 Transparency Toward Students

Although students are aware that the assistant operates within a controlled, course-specific environment shaped by the instructor's design choices, they are required at the beginning of each session to explicitly accept the platform's Terms and Conditions. This process ensures informed

consent regarding the use of AI within the course context. The Terms clearly outline user responsibilities, data usage policies, and the inherent limitations of generative AI systems.

Students are also constantly reminded via UI disclaimers that generative AI systems may produce inaccurate or hallucinated responses.

5 Institutional Governance Layer

5.1 Compliance with the AI Act

Under Annex III (b), AI systems intended to evaluate learning outcomes, particularly when such evaluation may steer educational trajectories, are classified as high-risk. In practical terms, a generative AI assistant integrated into a learning management system could easily fall within this category if it were allowed to assign numerical scores, determine pass/fail outcomes, or issue performance evaluations affecting academic standing. Even formative feedback mechanisms may generate interpretative ambiguity where they influence, directly or indirectly, the instructor's formal assessment.

Given the evolving nature of EU-level interpretative guidance regarding the boundary between formative support and formal evaluation, the institution adopted a precautionary and risk-mitigation-oriented approach, particularly considering the reputational and institutional implications of large-scale deployment.

To address this risk exposure, a dedicated institutional guardrail was embedded into the system architecture. This guardrail:

- Technically prevents the generation of grades, pass/fail determinations, or any evaluative outputs that could influence formal academic assessment.
- Implements additional language-based filtering, safeguards for ethically sensitive topics, and prompt-injection detection mechanisms to prevent circumvention attempts.

The guardrail is centrally defined at the institutional level and cannot be modified or overridden by individual instructors. For transparency purposes, instructors can review the guardrail configuration and understand the applied rules, although they cannot alter them.
This architectural decision positions the platform within a limited-risk configuration, in line with a compliance-oriented governance strategy.

5.2 Terms and Conditions

In addition to technical guardrails, the platform is governed by dedicated Terms & Conditions of Use developed in collaboration with the University's Data Protection Office and Legal Affairs team.

The document functions as a formal compliance layer and establishes:

1. **Strict educational scope limitation:** defines the Tool as exclusively dedicated to academic support within institutional courses, explicitly excluding private, commercial, or high-risk uses under the AI Act.
2. **Explicit prohibition of grading and automated decision-making:** formally prohibits the use of the Tool for grading, profiling, admission decisions, or any automated decision-making processes.

3. **Data protection**: establishes GDPR-aligned data governance principles, including prohibition of training on user data, data minimization, limited retention periods, and discouragement of entering personal or sensitive data.
4. **Allocation of responsibility**: clarifies that users remain fully responsible for how Outputs are interpreted and used, limiting institutional liability for misuse or misinterpretation.
5. **Intellectual property protection**: regulates ownership and permitted uses of course materials, Student Uploads, and Outputs.
6. **Monitoring mechanisms**: grants the University audit rights, log monitoring capabilities, and the authority to suspend access or initiate disciplinary procedures in cases of misuse.

5.3 Internal Monitoring

Alongside contractual and technical safeguards, the platform includes an internal monitoring system designed to supervise usage trends, monitor costs and ensure responsible large-scale deployment.

A dedicated institutional dashboard provides access to aggregated metrics such as overall interaction volumes, interaction frequency across courses, temporal usage patterns, operational costs, and the number and typology of blocked prompts.

Monitoring activities are conducted in compliance with data protection principles. Data are primarily processed in aggregated or anonymized form, and no profiling of individual users is performed for evaluative purposes. Access to identifiable interaction logs is strictly limited to cases of suspected misuse and governed by institutional data governance policies.

6 Adoption, User Feedback and Future Development

6.1 First Semester of Pilot Deployment

The platform was initially deployed in a controlled pilot phase involving a limited number of courses and instructors (7 courses, 13 classes). Among the 2,300 students enrolled in the participating courses, 24% actively engaged with the tool. Usage patterns indicate a progressive increase over time, with engagement peaks occurring at course launch and during exam preparation.

6.2 Students Surveys

In addition to usage analytics, a structured survey was conducted on a sample of approximately 100 voluntary students. The results provide evidence of the perceived pedagogical value of the Tool.

Although 35% report not using the Tool during the course, primarily due to habitual reliance on other external AI tools, the majority of active users express a clear preference for the contextual integration offered by the institutional assistant.

Across all major satisfaction indicators, the results show predominantly positive evaluations:

Indicator	Value	Interpretation
Active Usage Rate	65%	Two-thirds of students used the Tool at least once during the course.
Perceived Learning Impact	69%	Students reporting "A lot" or "Very much" improvement in learning.
Trust Level	67%	Students expressing high trust in the generated content.
Adoption Demand (Yes)	71%	Students who would like the Tool to be available in other courses.
Positive Overall Rating (4–5/5)	60%	Majority overall satisfaction despite pilot-stage limitations.
Preference vs. Commercial AI Tools	57%	Among users of external AI tools, a majority recommend the institutional assistant.

Table 1: Survey Results

84% already use commercial AI tools, yet 71% want the AI Course Assistant in other courses and 57% recommend it over other systems.

6.3 Future Developments: Agents

The next development phase focuses on enhancing AI assistants with more complex agent systems capable of structured interaction flows.

A first step in this direction is the introduction of a Study & Learn Mode Agent, designed to structure the study process in a more interactive and maieutic manner.

The Study Mode agent:

1. Lists available documents within the assistant's Knowledge Base.
2. Allows the student to select a document for revision.
3. Progressively analyzes the document, tracking the learning progress (e.g., number of processed sections).
4. For each section, the agent:
    a. Provides a concise explanation of key concepts.
    b. Generates flashcards.
    c. Proposes self-assessment quizzes (multiple choice, true/false, open-ended).
5. Delivers feedback on student responses.
6. Waits for the student's answer before proceeding.

This design introduces a structured, interactive learning loop:
explanation → self-assessment → feedback → progression.

To conclude, our ongoing goal is to integrate emerging market innovations selectively, ensuring that they respond to the real needs of instructors and learners and enhance meaningful pedagogical outcomes.

References

Ji et al., Z. (2023). *Survey of hallucination in natural language generation. ACM Computing Surveys.*

Jin et al., Y. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines,.

Lewis et al., P. (2020). Advances in Neural Information Processing Systems (NeurIPS 2020).

Qian, Y. (2025). Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field. TechTrends 69, 1105–1120.

Regulation (EU), A. (2024). *Annex III: High-Risk AI Systems referred to in Article 6 (2)*. Retrieved from https://artificialintelligenceact.eu/annex/3/.