Evaluating GenAI Innovation in Higher Education

A Whitepaper

Dr. Sandra Schön – https://orcid.org/0000-0003-0267-5215
Dr. Benedikt Brünner – https://orcid.org/0009-0005-8484-9160
Dr. Martin Ebner – https://orcid.org/0000-0001-5789-5296
Dr. Philipp Leitner – https://orcid.org/0000-0001-8883-6758

Graz University of Technology, Austria

Version 1.0 – May 22, 2026

Contact: martin.ebner@tugraz.at

Citation suggestion: Schön, S., Brünner, B., Ebner, M. & Leitner, P. (2026). Evaluating GenAI Innovation in Higher Education. A Whitepaper. Graz University of Technology. https://doi.org/10.3217/1tfkj-ntj85

Please note: A summary of the Whitepaper was accepted at EdMedia Conference 2026 as poster. The description of the poster presentation in EdMedia proceedings 2026 is partly identical with this text.

Abstract

The rapid integration of Generative Artificial Intelligence (GenAI) into higher education presents a dual challenge: leveraging its potential for personalized learning and content creation while mitigating risks related to data sovereignty, algorithmic bias, and pedagogical efficacy. Current discourse often lacks structured frameworks for systematically evaluating GenAI innovations beyond technical feasibility. This whitepaper addresses this gap by proposing a practice-derived, five-stage evaluation framework developed through a cross-case synthesis of different GenAI implementations at Graz University of Technology. Drawing on real-world use cases ranging from AI-generated multilingual educational videos and RAG-based chatbots to automated assessment tools, we outline a non-linear lifecycle approach comprising: (a) specifying context and use cases, (b) assessing feasibility (legal, ethical, and technical), (c) selecting implementation strategies, (d) piloting with multi-layered evaluation, and (e) performing data-informed analysis. The framework emphasizes that GenAI must be treated as a pedagogical intervention requiring continuous governance, human oversight, and institutional accountability rather than a static technological tool. We argue that sustainable adoption depends on the ability of institutions to iteratively refine, scale, or discontinue applications based on evidence of educational value and compliance. This work, already accepted and discussed as a poster presentation at the EdMedia Conference 2026, aims to provide higher education institutions with a robust, adaptable methodology for navigating the complexities of responsible GenAI innovation.

Key words: Generative AI, Higher Education, Evaluation Framework, Digital Sovereignty, Pedagogical Intervention, AI Governance

Content

1. Introduction

Generative Artificial Intelligence (GenAI) operates through a structured pipeline that transforms user intent into intelligent output, as shown in Fig. 1. This process begins with the query, in which the user provides a prompt that may be enriched with uploaded documents, prior chat history, contextual information and specific formatting instructions. This query is then mediated by the application layer, which serves as the gateway between the user and the underlying model. At this layer, the software environment, interface design, provider-defined system prompts, and embedded guidelines and constraints shape how the request is interpreted and processed.

The request is subsequently passed to the core GenAI Model, where learned architectural parameters, such as model weights, interact with adjustable configuration settings, including temperature for creativity, token limits for length, and other generation parameters. These elements directly influence the model's reasoning patterns, creativity, response length, and overall generation behavior. Finally, the system produces an output, delivering the generated response back to the user.

Understanding this flow, from the composition of the initial query to the configuration of the model and the constraints imposed by the application layer, is essential for evaluating how different inputs, settings, and usage contexts shape the quality, reliability, and behavior of the GenAI outputs.

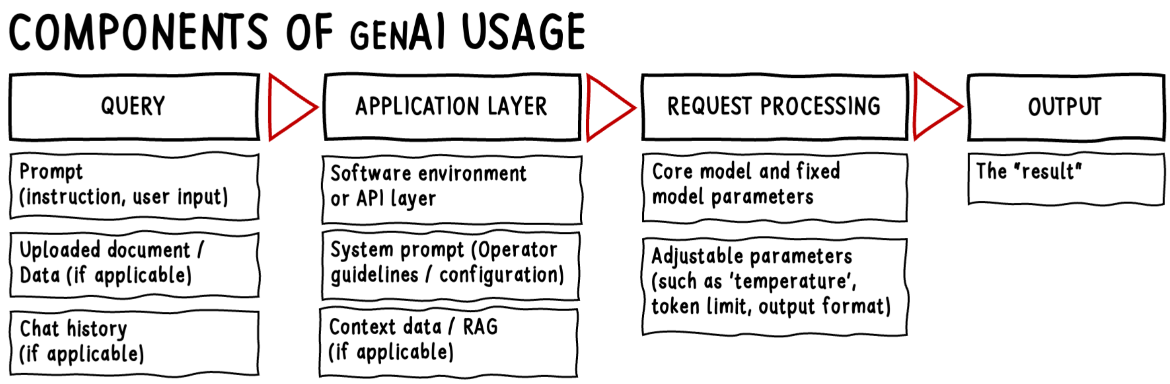

Figure 1: Components of GenAI usage

GenAI is no longer a peripheral educational technology. It increasingly affects how learning resources are produced, how students receive support, how assessments are generated, and how institutions organize teaching and learning (Liang et al., 2026; Nagler et al., 2025). AI-enabled systems promise personalization, adaptive feedback, multilingual access, accessibility support, and new forms of interaction with educational content.

At the same time, they raise concerns about privacy, bias, overreliance on automation, transparency, equity, academic integrity, authenticity, and the preservation of human agency in learning environments (Miao & Holmes, 2023; Van Woensel, 2025).

As illustrated in Fig. 2, GenAI in higher education encompasses a diverse spectrum of applications spanning instructional preparation, the development of educational resources, learner support mechanisms such as automated chat interfaces in online courses, grading and feedback facilitation, as well as learning analytics.

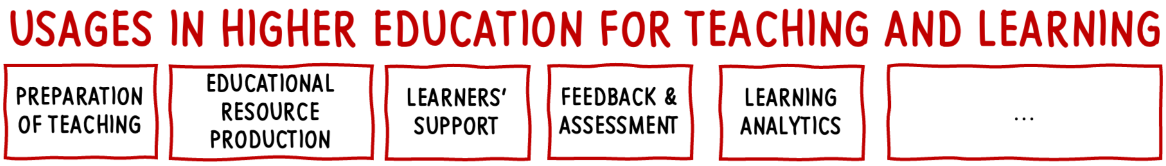

Figure 2: Usages of GenAI in Higher Education for Teaching and Learning

The adoption of generative AI in higher education involves navigating several interconnected aspects, which we have structured in Fig. 3 into five key areas for clarity. The deployment model ranges from commercial cloud services, offering rapid access but limited data control, to hosting on premise, which ensures full digital sovereignty and data ownership at the cost of higher infrastructure demands, with hybrid and sovereign cloud options bridging the gap. Open source developments are particularly interesting in this context, as they contribute significantly to digital sovereignty by reducing dependency on proprietary systems.These choices are directly linked to compliance and governance, where adherence to regional data privacy regulations - such as the European Union's General Data Protection Regulation (GDPR) - and strict data residency requirements dictate the feasibility of specific hosting arrangements. Complementing this is the architectural strategy, determining how the system interacts with institutional knowledge - whether through Retrieval-Augmented Generation (RAG) to ground responses without altering the base model, fine-tuning for domain specialization, or federated learning for privacy-preserving collaboration. The viability of these approaches further depends on operational readiness, encompassing the necessary Machine Learning Operations (MLOps) expertise, maintenance capabilities, and the financial balance between cloud-based operational expenditures and on-premises capital investments. Finally, successful implementation relies on system integration, ensuring that AI tools can interoperate with existing Learning Management Systems (LMS) like Moodle through standard APIs and authentication protocols to facilitate smooth adoption.

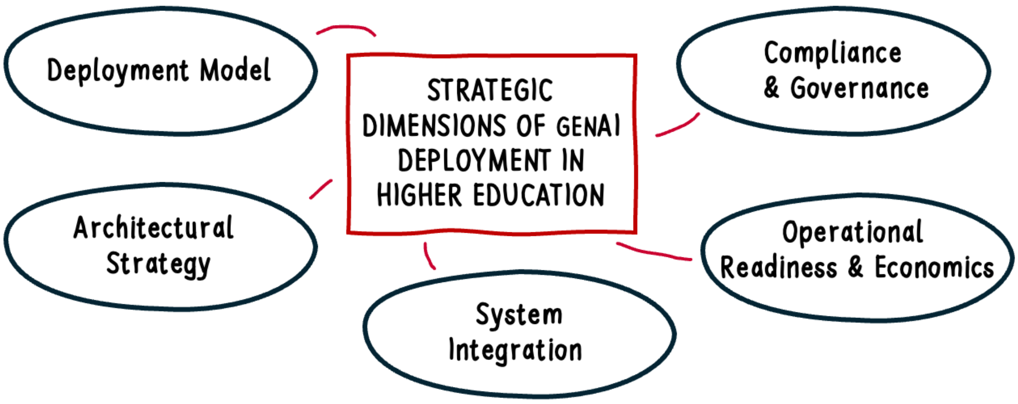

Figure 3: Strategic Dimensions of GenAI Deployment in Higher Education

Consequently, the central question for higher education institutions is no longer whether to explore GenAI, but how to establish systematic frameworks for its evaluation, governance, and continuous improvement, with particular relevance for learning and teaching (Brünner, 2025). Many GenAI initiatives begin as prototypes, pilots, or local experiments. However, sustainable implementation requires more than experimentation. It requires explicit processes for deciding whether an application is pedagogically useful, technically feasible, legally permissible, ethically defensible, secure, accessible, and institutionally scalable. It also requires the ability to decide when GenAI should not be used because a simpler, more reliable, or less risky solution would better serve the educational purpose.

This need is reinforced by the European regulatory and governance context. The EU Artificial Intelligence Act establishes a risk-based framework for trustworthy and human-centric AI, with attention to fundamental rights, transparency, safety, human oversight, robustness, and governance (European Parliament & Council of the European Union, 2024). For higher education, this is relevant because AI systems may affect access to learning, feedback, assessment, learner support, and educational decision-making. Systems used merely to support optional learning activities differ substantially from systems that influence assessment, progression, or institutional decisions. Therefore, legal and ethical feasibility must be assessed early rather than treated as a final compliance check.

GenAI evaluation is also challenging because GenAI systems differ from traditional educational technologies. Their outputs are probabilistic, context-sensitive, and often non-deterministic. Work on GenAI quality evaluation in software engineering shows that conventional quality assurance approaches are often insufficient and that practitioners rely on context-specific metrics, human review, automated checks, and AI-assisted evaluation methods (Yu et al., 2026). This insight transfers directly to education: the quality of GenAI applications depends on the educational context, target group, content domain, output format, available data, intended pedagogical function, and risk level.

For this reason, GenAI evaluation should be understood as a lifecycle process rather than as a single output check. The United States of America National Institute of Standards and Technology (NIST) AI Risk Management Framework describes AI risk management as an ongoing process involving governance, mapping, measurement, and management across the AI system lifecycle (Tabassi, 2023). The NIST Generative AI Profile further emphasizes that GenAI introduces or intensifies risks across design, development, deployment, operation, and decommissioning (Autio et al., 2024).

At Graz University of Technology (TU Graz) the Educational Technology unit and in particular the Future Learning, Analytics & AI for Teaching team is responsible for supporting the use of emerging technologies across all faculties in our institution. Over the past few years, we have supported numerous GenAI implementations in teaching contexts. Within this whitepaper, we aim to organize and reflect on these experiences to establish a foundation for more systematic evaluation practices.

The whitepaper therefore addresses the following guiding question:

How can higher education institutions systematically evaluate quality-assured GenAI applications in ways that integrate pedagogical quality, technical feasibility, legal compliance, ethical responsibility, data governance, accessibility, security, digital sovereignty, and institutional sustainability?

2. Methodology: Practice-based cross-case synthesis

The framework was derived from multiple GenAI implementations, pilot studies, prototype developments, and institutional reflections conducted within the Educational Technology context at TU Graz. The goal was not to evaluate one specific tool, but to identify recurring decision points, risks, evaluation criteria, quality assurance mechanisms, and implementation pathways across different types of GenAI applications (see Table 1).

Table 1: Use Cases of Graz University of Technology, Educational Technolgy team (TU Graz)

Use Case	GenAI function	Main evaluation focus
Societech MOOC (Schön et al., 2025)	AI-generated course materials, videos, and self-assessments	Content quality, production effort, OER compatibility
OER in HE MOOC (Schön et al., 2025)	Generating multilingual human avatar-based videos	Translation, authenticity, consent, multilinguality
RMVPro / Rapid Multilingual Video Production (Brünner & Ebner, 2025)	AI-supported rapid production of multilingual educational videos with avatar presenters	Translation quality, comprehensibility, native-speaker feedback, authenticity, learner perception, and production efficiency
Synthetic Educators (Struger et al., 2025)	AI-generated teaching avatar	Recall, emotional response, perceived quality
QUEST (Ebner et al., 2025)	AI-generated MCQs	Assessment quality, distractors, clarity, cognitive demand
MOOC Chatbot SRL (Brünner et al., 2026a)	RAG-based learner support	Student interactions, SRL behavior, misuse
aicast.fyi (Brünner et al., 2025)	Personalized educational podcasts	Hybrid human-AI content, personalization, transparency
prompting.school (Brünner et al., 2026b)	AI-supported guided prompt engineering learning platform	AI literacy, prompt quality, learner engagement, SRL behavior

Building upon the analysis of the presented cases, the authors engaged in a reflective discussion to distill key insights for future initiatives. These insights served as the foundational basis for sketching a preliminary framework designed to guide subsequent projects. Complementing the analysis of these cases, we undertook supplementary literature research that was not conducted systematically. Relevant findings from this research were incorporated in the framework and address specifically methods for evaluating GenAI results.

3. Cross-case synthesis assumptions for the framework

The cases in Table 1 show that GenAI evaluation in higher education cannot rely on one universal metric. The Societech MOOC and OER in HE MOOC cases indicate that GenAI can accelerate the production of educational materials and multilingual resources, but also that expert review, legal checks, transparency, and instructional design remain necessary (Schön et al., 2025). The RMVPro case adds a more process-oriented perspective on multilingual video production. It shows that AI-supported workflows can substantially accelerate production, but require quality assurance for translation accuracy, cultural adequacy, pronunciation, terminology, and learner trust in AI avatars (Brünner & Ebner, 2025). The Synthetic Educators case demonstrates that technical realism is not sufficient as a quality criterion; learner perception, emotional response, authenticity, and disclosure are also important (Struger et al., 2025). The QUEST case shows that AI-generated assessment items require explicit quality criteria, including clarity, distractor plausibility, ambiguity, alignment with learning objectives, and cognitive demand (Ebner et al., 2025). The MOOC Chatbot SRL case illustrates that actual learner interaction may differ from the intended pedagogical use, and that log data and learning analytics are necessary to understand real use (Brünner et al., 2026a). The aicast.fyi and prompting.school cases show that personalization and AI literacy require attention to learner agency, transparency, and guided interaction with AI systems (Brünner et al., 2025; Brünner et al., 2026b).

Across the use cases, five recurring lessons emerge that we take as assumption for the framework development:

1. **GenAI should be evaluated as a pedagogical intervention, not merely as a technical tool.** A GenAI output may be fluent, visually convincing, or technically functional while still being pedagogically weak. An AI-generated multiple-choice question may contain weak distractors, a chatbot may reduce productive struggle, and a synthetic avatar may be understandable but not trusted by learners. Therefore, educational value depends on alignment with learning objectives, learner needs, instructional design, and the role of human teaching.
2. **Quality-assured GenAI depends on structured human-AI collaboration.** Across the cases, human expertise remained essential. Experts reviewed generated course materials, native speakers checked multilingual videos, subject-matter experts evaluated assessment items, and teachers or researchers interpreted learner interactions. GenAI can accelerate production, transformation, and support processes, but it shifts work toward planning, review, moderation, and evaluation rather than removing human responsibility.
3. **Different GenAI applications require different quality criteria.** Videos, quizzes, chatbot responses, podcasts, and prompt-engineering platforms cannot be evaluated with the same metric. Multilingual videos require attention to translation, pronunciation, culture, authenticity, and comprehensibility. Multiple-choice questions require attention to distractor quality and ambiguity. RAG-based chatbots require retrieval quality, grounding, context integrity, and interaction analysis. AI literacy platforms require attention to prompt quality, reflection, and self-regulated learning (Brünner et al., 2026c). This supports a flexible framework with general stages but use-case-specific evaluation criteria.
4. **Real use must be studied after deployment or pilot testing.** Several cases show that pre-deployment quality checks are necessary but insufficient. Learners may use a chatbot differently than intended, may understand AI-generated videos but still prefer human presenters, or may need guidance to use GenAI critically. Evidence from real use, such as log data, surveys, focus groups, teacher observations, and learning analytics, is

therefore necessary to understand educational value, misuse, acceptance, and unintended consequences.

5. **GenAI evaluation must be embedded in institutional governance.** Data protection, copyright, digital sovereignty, accessibility, security, and ethical responsibility cannot be left to individual project teams alone (see Andrews et al., 2022). To mitigate copyright infringements, strategies for the production and use of Open Educational Resources (OER) are needed (Schön & Ebner, 2025). Institutional work on AI, Learning Analytics, and digital ethics highlights the need for explicit roles, procedures, and accountability structures in higher education (Ebner et al., 2022). This is particularly important because GenAI systems are not static. Their behavior may change when models are updated, prompts are revised, providers modify terms, or source documents are changed. Quality assurance therefore requires not only development and evaluation, but also monitoring, revision, and the possibility of stopping or retiring a system.

4. A framework evaluating GenAI innovations for higher education

The resulting suggested framework should therefore be understood as a practice-derived model for structured institutional evaluation. Its purpose is to support researchers, EdTech developers, instructional designers, teachers, legal and data protection experts, AI governance bodies, and institutional decision-makers in evaluating GenAI applications before implementation, scaling, or long-term operation.

Core elements of the framework are shown in Fig. 4. Our proposed framework outlines a dynamic, five-stage cycle for evaluating GenAI in higher education: (a) Specifying Context and Use Cases, (b) Assessing Feasibility, (c) Selecting Implementation Strategies, (d) Piloting and Conducting Multi-Layered Evaluation, and (e) Performing Data-Informed Analysis. Crucially, this process is designed to be non-linear and iterative. Rather than following a rigid sequence, each phase concludes with a critical Sustain-or-Discontinue decision point. If pre-defined criteria are not met, the workflow loops back to a preceding phase for refinement, or the initiative may be discontinued. This iterative decision-making mechanism supports the evidence-informed allocation of institutional resources towards GenAI innovations with demonstrated viability and potential impact, while reducing the risk of scaling ineffective or misaligned solutions.

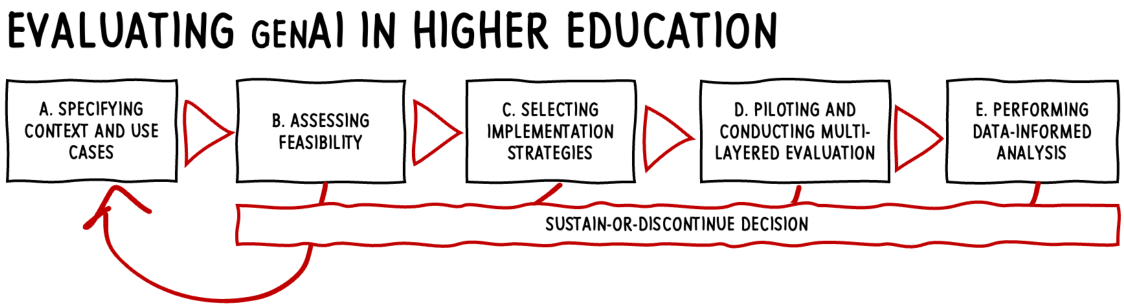

Figure 4: Core processes for the evaluation of GenAI in higher education

A. Specifying context and use cases

The process initiates with a request for educational technology, which must be systematically refined through feedback from subsequent stages. Crucially, the use case should be formulated as an educational problem to be solved, rather than a mere desire to utilize a specific tool. Teachers, instructional designers, or project initiators must clarify the core problem, the affected stakeholders, the anticipated educational benefit, and whether GenAI is strictly necessary. If a simpler, more reliable, or lower-risk solution, such as a glossary, FAQ, search function, or decision tree, can address the issue, GenAI should not be the default choice.

This phase encompasses use-case specification and pedagogical alignment. The team must define learning objectives, the target group, the educational setting, expected interactions, output formats, human responsibilities, and system boundaries. Equally important is defining what the system must not do (e.g., providing direct solutions to graded assignments or replacing teacher feedback). The central inquiry is whether the proposed application supports the intended learning activity without undermining learner agency, self-regulation, or academic integrity.

Furthermore, the institution must identify the project owner, contributing experts, decision-making authorities, expected costs, required maintenance, and long-term accountability. The team should then classify the use case according to risk, considering the degree of automation, the handling of personal data, the relation to assessment, potential learner harm, and the level of human oversight. This aligns with risk-based AI governance approaches, such as the EU Artificial Intelligence Act and broader AI risk-management frameworks (Autio et al., 2024; European Parliament & Council of the European Union, 2024; Tabassi, 2023).

B. Assessing feasibility

This phase entails a comprehensive assessment covering data, legal, ethical, security, and copyright dimensions. The evaluation must extend beyond simple server location to encompass the entire data processing chain: including providers, subcontractors, data retention policies, model-improvement settings, access to logs, deletion procedures, contractual safeguards, and auditability.

If personal data, student submissions, learning analytics, or institutional documents are processed, data protection requirements must be assessed on a case-by-case basis (European Data Protection Board, 2024). Additionally, the team must rigorously consider copyright issues, consent mechanisms, transparency requirements, accessibility standards, risks of prompt injection, data leakage, potential misuse, and broader ethical implications before any implementation proceeds.

Open-licensed materials are particularly well-suited for training and utilizing GenAI, as their clear permissions reduce legal uncertainties regarding data usage. Therefore, it is highly beneficial for universities to either develop their own tailored Open Educational Resources (OER) or leverage existing OER from other institutions to support these AI initiatives (Schön & Ebner, 2025).

C. Selecting implementation strategies

This stage focuses on implementation-pathway selection. The institution must decide between three primary paths: utilizing an existing tool, adapting/configuring an existing tool, or developing a custom solution.

- Existing tools may suffice for low-risk use cases where human review is guaranteed.
- Adapted tools are appropriate when course-specific grounding, controlled access, or specific institutional configurations are required.
- Custom solutions become necessary when pedagogical, privacy, sovereignty, accessibility, or integration requirements cannot be met by off-the-shelf alternatives.

D. Piloting and conducting multi-layered evaluation

This phase combines prototype design and implementation with a formative and summative evaluation design. The selected solution is built or configured with meticulous attention to prompt engineering, model selection, interface design, source-document preparation, RAG where appropriate, access control, moderation, transparency disclosures, logging, evaluation setups, and fallback procedures. The objective is not merely to create a technically functional prototype, but one capable of being systematically evaluated.

Prior to learner interaction, the team must define acceptance criteria, test cases, rubrics, reviewer roles, and documentation procedures for assessing GenAI's quality for a specific purpose (see Fig. 5). Evaluation protocols should include typical, difficult, ambiguous, and adversarial inputs. Depending on the specific use case, the assessment must cover factual correctness, pedagogical usefulness, accessibility, bias, robustness, safety, citation quality, format compliance, usability, and alignment with task boundaries. While AI-assisted review can support the process, it must never replace expert human judgment.

As shown in Fig. 5, to accurately assess GenAI performance, one must systematically describe and evaluate several interconnected factors. Therefore, the inquiry context requires a detailed definition of the domain, specific use case, and complexity level, alongside a description of iterative interaction strategies, attached resources, used language, and (if) repetition of requests. Then, for processing with GenAI, the right tool(s) must be identified to be used for the inquiry. The application layer, see Fig. 1, must be programmed or configured and examined to understand how the software environment and provider-defined system prompts influence the interaction. Additionally, the assessment involves the core GenAI model configuration, fixed architectural parameters, and adjustable runtime settings (like temperature) that drive the output. Finally, a robust evaluation framework for the results must integrate automated metrics, human judgment, and user feedback, while explicitly accounting for the unique characteristics and variations across different systems.

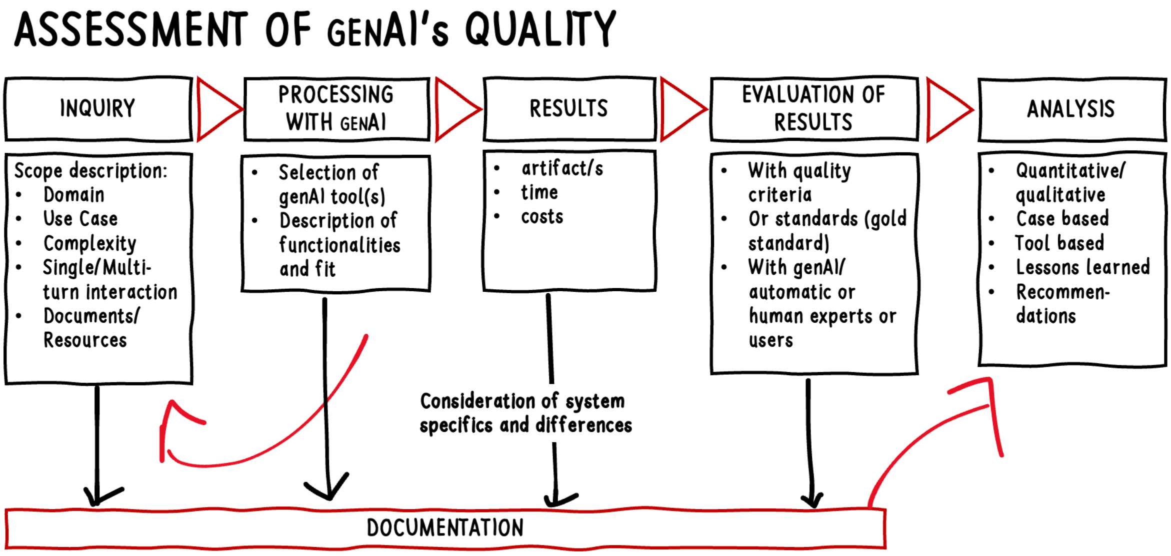

Figure 5: Assessment approaches for GenAI results and usage

Following the prototype and evaluation phase, a controlled pilot and deployment is mandatory. A pilot precedes full-scale deployment to observe whether users understand the system, if the tool integrates meaningfully into learning activities, if learners exhibit over-reliance, if teachers can facilitate it effectively, and if unforeseen risks emerge.

E. Performing data-informed analysis

The final stage focuses on monitoring, evidence collection, and institutional learning. Post-deployment, the institution must gather diverse evidence, including log data, user feedback, teacher observations, surveys, focus groups, error reports, misuse attempts, and relevant learning analytics. Crucially, high usage rates alone must not be equated with educational value.

The final analysis determines whether the application should be continued, revised, scaled, restricted, or retired. Lessons learned must feed back into future GenAI projects, staff training programs, governance policies, evaluation rubrics, procurement decisions, and AI literacy initiatives.

Broad deployment should only occur after a successful piloting phase and upon confirmation of positive analysis results. Full deployment requires transparency, user guidance, support channels, reporting mechanisms, and clear rules regarding academic integrity and appropriate use.

Overarching principle: The lifecycle approach

Overall, this process treats GenAI evaluation as a continuous lifecycle rather than a one-time output check. A GenAI application is deemed appropriate only if it is educationally meaningful, legally and ethically acceptable, technically robust, accessible, secure, understandable to users, and maintainable by the institution. The framework includes an implicit revision, redesign, or stop decision at the end of the lifecycle. If evaluation reveals insufficient quality, the project must return to an earlier stage. Revisions may involve narrowing the use case, refining prompts or models, enhancing the knowledge base, increasing human review, strengthening moderation, improving accessibility, or altering the implementation pathway. In some instances, terminating the project is the most responsible and ethical decision.

5. Recommendations for higher education institutions

Higher education institutions should treat GenAI applications as pedagogical interventions rather than as isolated tools. Each project should begin with an educational problem and an AI added-value check. The first institutional question should be whether GenAI provides meaningful educational value compared with simpler, more stable, or less risky alternatives.

- Institutions should specify the use case before selecting tools. Learning objectives, target group, output format, data sources, intended interaction, risk level, and human responsibilities should be clarified before implementation. This prevents tool-driven adoption and makes later evaluation more precise.
- Institutions should classify risk early. Applications that support optional learning activities require a different level of governance than applications that influence assessment, progression, learner profiling, or educational decision-making. Risk classification should determine the required level of human oversight, documentation, legal review, monitoring, and institutional approval.
- Institutions should combine expertise. Effective GenAI evaluation requires collaboration among teachers, subject-matter experts, EdTech experts, instructional designers, developers, data protection officers, open data and OER experts, legal experts, accessibility specialists, security experts, ethics experts, and users. No single role can responsibly assess all dimensions of a GenAI application.
- Institutions should separate prototyping, evaluation, pilot testing, and deployment. A working prototype is not yet a deployable educational system. GenAI applications should pass structured evaluation and, where appropriate, controlled pilot testing before being scaled to broader educational use.
- Finally, institutions should monitor GenAI systems after deployment and should establish feedback loops. Monitoring should include user feedback, error reporting, interaction data, documentation of system changes, and clear procedures for pausing, revising, or retiring applications. Responsible GenAI use requires not only implementation, but also continuous institutional learning.

6. Preliminary conclusion

GenAI innovation in higher education requires more than experimentation. It requires structured, evidence-informed, and ethically grounded evaluation practices. The use cases presented in this white paper show that GenAI can support educational content production, multilingual access, assessment generation, learner support, personalization, AI literacy development, and institutional innovation. At the same time, they demonstrate recurring risks: superficial content, weak assessment items, unclear authenticity, learner ambivalence toward synthetic media, misuse attempts, data governance challenges, accessibility issues, and insufficient alignment with pedagogical goals.

The proposed framework makes institutional decision-making explicit. It guides higher education institutions from an initial educational request through specification, responsibility assessment, risk classification, governance review, implementation, evaluation, revision, pilot testing, deployment, monitoring, evidence collection, and final analysis. The framework emphasizes that GenAI quality cannot be reduced to technical performance or efficiency. It must be assessed in relation to educational value, learner agency, human oversight, legal and ethical acceptability, data governance, digital sovereignty, and long-term maintainability.

The central institutional capability is therefore not the ability to adopt GenAI quickly. It is the ability to evaluate GenAI responsibly, to learn from its use, to revise applications when necessary, and to stop or retire systems that do not meet educational, legal, ethical, or institutional requirements. In this sense, responsible GenAI innovation includes not only the courage to experiment, but also the discipline to govern, monitor, and discontinue.

7. Limitations and future work

This framework is derived from use cases at TU Graz and should be adapted before transfer to other institutional, legal, cultural, and pedagogical environments. Although the cases cover a wide range of GenAI applications, they do not represent all possible higher education contexts. Institutions with different governance structures, legal obligations, technological infrastructures, or educational traditions may need to modify the process.

The included cases also differ in maturity and methodological rigor. Some are empirical studies, others are prototype developments, pilot reports, technical evaluations, or governance reflections. Therefore, the framework should be understood as a practice-derived process model rather than as a validated maturity model, standardized assessment instrument, or universal quality metric.

Future work should test the process across additional universities, disciplines, and regulatory contexts. It should also develop more specific rubrics for different GenAI output types, including chatbot responses, synthetic videos, AI-generated assessment items, personalized media, RAG systems, AI literacy platforms, and multilingual educational content. Further research should examine how institutions can operationalize risk classification, digital sovereignty, accessibility, security testing, monitoring, and retirement decisions.

Disclaimer on AI usage

Based on Academic Integrity and Transparency in AI-assisted Research and Specification Framework (Bozkurt, 2024), the authors of this whitepaper acknowledge that it was drafted with the assistance of Lumo, ChatGPT 5.5, and DeepL (Versions as of May 2026), complementing the human editorial process. The human authors critically assessed and validated the content to maintain academic rigor. The authors also assessed and addressed potential biases inherent in the AI-generated content. The final version of the paper is the sole responsibility of the human authors.

References

Andrews, D., Leitner, P., Schön, S., Ebner, M. (2022). Developing a Professional Profile of a Digital Ethics Officer in an Educational Technology Unit in Higher Education. In: Zaphiris, P., Ioannou, A. (eds) Learning and Collaboration Technologies. Designing the Learner and Teacher Experience. HCII 2022. Lecture Notes in Computer Science, vol 13328. Springer, Cham. https://doi.org/10.1007/978-3-031-05657-4_12

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Bozkurt, A. (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. Open Praxis, 16(1), 1-10. https://doi.org/10.55982/openpraxis.16.1.654

Brünner, B. (2025). Presentation: From users to creators: How universities can prepare students for *the future of AI*. [Original German: Präsentation: Von Nutzer:innen zu Gestalter:innen: Wie Hochschulen auf die KI-Zukunft vorbereiten können] Graz University of Technology. https://doi.org/10.3217/9vg04-95q04

Brünner, B., Deutschmann, F., Etzelstorfer, S., Lechner, A., Schön, S., & Ebner, M. (2026c). AI-powered chatbots for education using RAGs: A study on embedding efficiency, LLM performance, and context integrity. In S. Uğur (Ed.), *Transforming education with singularity technologies: Lifelong learning from childhood to adulthood* (1st ed., Chapter 9). Chapman and Hall/CRC. https://doi.org/10.1201/9781003584339

Brünner, B., & Ebner, M. (2025). Rapid multilingual video production utilizing artificial intelligence. In T. Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning 2025* (pp. 1223–1230). Barcelona, Spain: Association for the Advancement of Computing in Education. https://www.learntechlib.org/p/226280

Brünner, B., Ebner, M., & Schön, S. (2026a). Exploring GenAI chatbots in MOOCs: Analyzing student interactions and self-regulated learning behaviors. In E. Hamonic & R. Sharrock (Eds.), *Digital education: Shaping sustainable lifelong learning for all in the era of AI* (Lecture Notes in Computer Science, Vol. 15733). Springer. https://doi.org/10.1007/978-3-032-00056-9_2

Brünner, B., Jahic, I., & Ebner, M. (2026b). Prompting.school: A design-based research approach to teaching AI literacy through guided prompt engineering practice. In M. E. Auer & P. Toth (Eds.), *Innovation via collaborative learning in engineering education* (Lecture Notes in Networks and Systems, Vol. 1848). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-20381-6_62

Brünner, B., Leitner, P., Geier, G., & Ebner, M. (2025). aicast: Combining AI-generated and instructor-defined content in educational podcasts. In T. Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning 2025* (pp. 160–164). Barcelona, Spain: Association for the Advancement of Computing in Education. https://www.learntechlib.org/primary/p/226139/

Ebner, M., Brünner, B., Forjan, N., & Schön, S. (2025). Ensuring quality in AI-generated multiple-choice questions for higher education with the QUEST framework. In Ł. Tomczyk (Ed.), *New media pedagogy: Research trends, methodological challenges, and successful implementations* (Communications in Computer and Information Science, Vol. 2537). Springer. https://doi.org/10.1007/978-3-031-95627-0_20

Ebner, M., Schön, S., Andrews, D., Leitner, P., & Thurner, S. (2022). *Artificial intelligence and learning analytics as a challenge: Tasks of ethics officers and ethical requirements for the use of digital technologies in higher education [Original German: Künstliche Intelligenz und Learning Analytics als Herausforderung: Aufgaben von Ethikverantwortlichen und ethische Anforderungen beim Einsatz digitaler Technologien in der Hochschullehre]* (Preprint). Graz University of Technology. https://doi.org/10.3217/3ej3a-j9j33

European Data Protection Board. (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*. https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689*. http://data.europa.eu/eli/reg/2024/1689/oj

Liang, Z., Yang, K., Sha, L., Gašević, D., Yan, L., & Chen, G. (2026). A systematic review of generative AI in education: Empirical insights from a human–AI interaction perspective. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.70055

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Nagler, W., Mair, B., Ebner, M., & Schön, S. (2025). Inspired by generative AI, ignoring open knowledge? Findings from a longitudinal study on digital equipment and activities of study beginners at TU Graz (2011–2024). In T. Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning* (pp. 1317–1325). Barcelona, Spain: Association for the Advancement of Computing in Education. https://www.learntechlib.org/p/226272

Schön, S., Brünner, B., Ebner, M., Edelsbrunner, S., Hohla-Sejkora, K., & Uhl, B. (2025). Early findings from pilots in AI-driven education: Effects of AI-generated courses and videos on learning and teaching. In M. E. Auer & D. May (Eds.), 2024 Yearbook *Emerging Technologies in Learning* (Learning and Analytics in Intelligent Systems, Vol. 44). Springer. https://doi.org/10.1007/978-3-031-80388-8_2

Struger, P., Brünner, B., & Ebner, M. (2025). Synthetic educators: Analyzing AI-driven avatars in digital learning environments. In B. K. Smith & M. Borge (Eds.), *Learning and collaboration technologies* (Lecture Notes in Computer Science, Vol. 15807). Springer. https://doi.org/10.1007/978-3-031-93567-1_13

Schön, S., & Ebner, M. (2025). Beyond Open Access: Open Educational Resources for Legal Clarity, Sustainability, and Digital Sovereignty in European University Alliances. Weizenbaum Journal of the Digital Society, 5(4). https://doi.org/10.34669/wi.wjds/5.4.2

Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Van Woensel, T. (2025). *A white paper on AI and education*. Eindhoven University of Technology. https://tomvanwoensel.com/wp-content/uploads/2025/04/ai-white-paper-van-woensel.pdf (last access 2026-04-29)

Yu, L., Alégroth, E., Chatzipetrou, P., & Gorschek, T. (2026). Evaluating the quality of GenAI applications in software engineering: A multi-case study. *Empirical Software Engineering, 31*, Article 29. https://doi.org/10.1007/s10664-025-10759-2