Rethinking University Learning: Course-Based Human–AI Interaction in a Controlled Educational Environment

Miklós Szabó[1], Renáta Németh[2] and Annamária Tátrai[3]

[1]ELTE Eötvös Loránd University, Faculty of Social Sciences, Department of Minority Studies
[2]ELTE Eötvös Loránd University, Faculty of Social Sciences, Department of Statistics, Research Center for Computational Social Science
[3]ELTE Eötvös Loránd University, Faculty of Social Sciences, Department of Statistics, Data for Good Research Group

Abstract

This study examines how generative AI can be integrated and interpreted in higher education through a controlled research environment. Drawing on interaction logs, ethnographic data, and focus group discussions, it analyses how students and instructors adapt the technology within their pedagogical contexts. Rather than replacing human teaching, AI functions as a reflective medium that reveals institutional assumptions and reshapes relations of trust, authority, and learning.

Keywords: AI in education; anthropology of technology; ethnography of learning.

1. Introduction

In the past four years, artificial intelligence has become an everyday experience rather than merely an engineering solution, a scientific problem, or a science-fiction concept. This transformation has also brought a profound shift in how we contextualize technology. Public and academic discourse has moved away from asking what AI objectively or technologically is, toward examining how "we," the users, make sense of it. The social implementation of a technology, and the consequences of its integration, are shaped to a great extent by the worlds we imagine as possible through it, often more so than by its actual technical operation. As Solomon Asch observed, "Most social acts have to be understood in their setting, and lose meaning if isolated. No error in thinking about social facts is more serious than failure to see their place and function" (1952: 61). Inspired by this perspective and by the anthropological tradition of examining technologies within their social and cultural contexts, we initiated a research project in 2023 in response to the emergence and widespread availability of generative AI tools such as ChatGPT. Our preliminary institutional ethnographic inquiry conducted in this context identified three broad orientations toward generative AI in higher education. The first can be described as non-use, involving institutional bans or restrictions that prevent students from developing the skills necessary to integrate new modes of working into their academic and professional routines, potentially placing them at a disadvantage on the labour market. Also sets up the environment for the second: misuse, which refers to the "uneducated" application of the technology, limiting its potential to enhance productivity and fostering mistrust, leading institutions to adopt "catch and punish" approaches rather than constructive integration. The third, constructive use, represents a more technologically optimistic stance, in which institutions integrate AI in a structured way, ensuring equal access while using the curriculum to educate students about both its benefits and its inherent risks. These orientations provided the conceptual lens for designing our research environment. To examine these orientations in practice and to operationalize our anthropological perspective, we designed a research setting that would move beyond surface-level observation of AI use. A key challenge was the "black box" nature of large language models, which makes it difficult to analyse how interactions unfold and how information is processed. The implementation of a pre-existing commercial retrieval-augmented generation (RAG) system within the educational environment. The system allows control over the corpus and enables the observation of AI–human interactions in a controlled and transparent way. This controlled research environment enabled the collection of data at multiple levels: the data used as the course corpus, how the system itself mobilizes information to respond to user queries, and the actual student engagement with a large language model agent. This data can be organised and analysed to identify best practices and emerging trends, and to develop methodological insights relevant to the implementation of generative AI solutions in education more broadly. Crucially, the tool also functions as a research stimulus, providing contextualised interactional data for observing institutional, educator, and student responses through focus groups, institutional ethnography, and surveys.

2. System Design and Pedagogical Framing

The core of the system is a course-specific retrieval-augmented generation (RAG) interface built on a commercial large language model (LLM) platform. Instructors create a course platform, upload selected reading materials, define prompt structures, and activate the system for student use via a web-based chat interface. Unlike general-purpose models, the assistant is restricted to responses drawn from the approved corpus, which typically includes textbooks, lecture slides, and other course readings. This design supports pedagogical control and traceability while allowing students to engage dynamically with course content. The technical implementation has been described in detail in our previous publications (Németh et al., 2024; Németh et al., 2025); here we highlight the elements most relevant to the present deployment. Technically, the system is built on a layered architecture. The ingestion layer processes course materials through segmentation and vectorisation, storing them in a secure retrieval database. The query layer identifies and matches student questions with the most relevant content segments. The response layer then employs the language model to generate answers that remain constrained by the retrieved material. Finally, the logging layer records all interactions along with their metadata. This architecture ensures that every response is grounded in the original course content while also producing structured data for subsequent pedagogical and interaction analysis. This design reflects two key principles. First, retrieval grounding and explicit referencing aim to mitigate hallucination and build student trust, aligning with current work in explainable AI and educational design research (Reeves, 2006). Second, systematic logging transforms the AI assistant from digital educational support into a research tool, providing a structured environment for observing how students and teachers interact with curricular content through AI mediation. Allowing that once interaction patterns are well understood, the tested pedagogical methods can be implemented with any LLM solution without significant deviation in use.

3. Methods

The ongoing research to date spans two academic years (2023–2025) and combines the controlled deployment of a retrieval-augmented generation (RAG) system with ethnographically informed inquiry. The aim was not only to document the system's technical operation but to understand how its introduction reshaped educational routines, pedagogical relations, and institutional interpretations. A mixed-methods approach (Creswell Plano Clark, 2018) integrated quantitative interaction data with qualitative material from focus groups, classroom observations, and institutional ethnography. The methodological framing drew on digital ethnography (Pink et al. 2016) and hypermedia ethnography (Dicks et al. 2005), which emphasise situating digital practices in their broader social and institutional contexts. Quantitative data came from system logs and instructor evaluations. Each student–AI interaction was logged with timestamps, course identifiers, and the passages retrieved to inform the assistant's responses. These data allowed us to analyse usage patterns, engagement with required readings, and interaction types, including temporal rhythms and degrees of conceptual synthesis (Németh et al. 2024, 2025). Instructor evaluations of stratified random samples of student–assistant exchanges were used to assess factual accuracy and response quality. Focus groups and ethnographic fieldwork complemented these quantitative sources. Mid- and post-semester focus groups with students and instructors explored perceptions of trust, control, and pedagogical usefulness, revealing interpretive dynamics not visible in log data. All data collection complied with GDPR and institutional ethical guidelines. Logs were anonymised at the point of collection, and participation in qualitative activities was voluntary and based on informed consent.

4. Deployment and Use Cases

Building on the methodological framework described above, this section provides an overview of the contexts in which the system was deployed, illustrating how disciplinary settings and pedagogical structures shaped its use. Since the autumn semester of 2023, the system has been deployed at thirteen Hungarian universities across a wide range of disciplinary contexts. More than 50 instructors and over 2,000 students have used the platform in courses spanning the social sciences, humanities, teacher education, law, and engineering. Over the past two years, the accumulation of recorded AI-human interactions has enabled further multidisciplinary research on the existing data.

5. Observations on Student Behavior

This section presents an interpretation of the data gathered during the pilot project, complemented by qualitative insights derived from focus group discussions with students from the same courses. A detailed quantitative analysis of the data collected during this period has been published in Németh et al. (2024). The introduction of the assistant did not produce a uniform mode of engagement. Instead, students and instructors incorporated it into existing pedagogical rhythms, disciplinary expectations, and interpretive practices in diverse ways. Log data and ethnographic observations revealed that the system became a site where students' learning strategies, trust negotiations, and disciplinary epistemologies intersected. Rather than replacing reading, teaching, or classroom dialogue, the assistant mediated new configurations of these practices.

5.1. Use Patterns

Students' use of the assistant followed distinctive temporal and interactional rhythms. Across courses, activity levels were shaped by the academic calendar, with clear spikes around midterm and final assignments. In cultural anthropology courses, the majority of queries occurred in the weeks leading up to essay deadlines, whereas in statistics preparation courses, interaction was distributed more evenly throughout the semester, reflecting a steadier rhythm of practice and review. This divergence extended to the form of interaction. In anthropology courses, roughly one-third of questions explicitly referenced assigned readings, and around fifteen percent displayed some degree of conceptual synthesis, such as comparing theoretical perspectives or asking for reformulations of key arguments. In statistics and network technologies courses, by contrast, interactions were dominated by formula checks, definitional clarifications, and problem-solving steps. These quantitative patterns were mirrored in students' reflections. Some anthropology students described using the assistant intensively only when writing essays, "diving into the texts" through it; others used it steadily to clarify readings during the semester, explaining that "it's just faster than flipping through all the readings again." The assistant thus became embedded within pre-existing learning strategies. For some, it served as an ongoing study partner; for others, as a strategic tool activated during moments of academic pressure.

5.2. Trust and Mediation

Trust in the assistant emerged through situated interactions rather than being assumed in advance. Students consistently pointed to referential transparency—the system's ability to ground answers in specific readings and page numbers—as crucial for establishing its credibility. Many explicitly contrasted this with their experiences of public LLMs, saying they "wouldn't believe it if it were just ChatGPT," but felt reassured when they could trace responses back to familiar texts. This anchoring enabled students to position the assistant within the normative framework of academic reading and interpretation, rather than as an external authority. Equally important was how students framed the role of the assistant in their learning. As a focus group discussion has shown, students tend to position AI systems as peers, tutors, or tools, and this framing shapes how trust is negotiated. In anthropology courses, students frequently treated the assistant as a knowledgeable peer or patient tutor, engaging in multi-turn dialogues to clarify concepts, reinterpret passages, or challenge inconsistencies. Questions like "Can you explain Mauss again, but in different words? I think I got lost the first time," or "Earlier you said that Sumner wasn't racist, but now you mention social Darwinism — which one is it?" exemplify this dialogic orientation. One student described the experience as "making the text speak back," highlighting how the assistant mediated between student and text rather than substituting for either. By contrast, students in statistics and network technologies courses typically positioned the assistant as a tool—a stable, impersonal reference source. Their questions were narrowly defined, often single-turn, and evaluated primarily on factual correctness. Trust, in other words, was not monolithic; it was enacted through role framing, disciplinary expectations, and interactional modes.

5.3. Disciplinary Differences

The most striking patterns emerged along disciplinary lines. In interpretive fields such as anthropology, the assistant became woven into students' dialogic engagement with texts. Students used it to negotiate meaning, compare perspectives, and clarify theoretical nuances. They frequently engaged in extended, iterative exchanges, evaluating trust through interpretive alignment and textual resonance. In technical fields like statistics and network technologies, the assistant functioned primarily as a structured reference instrument, facilitating efficient access to correct answers rather than interpretive dialogue. Here, trust was evaluated on the basis of correctness and reliability, and interactions remained short and instrumental. These disciplinary contrasts reflect underlying epistemic cultures rather than differences in technological affordances. In anthropology, the assistant was framed as a peer or tutor and integrated into interpretive practices; in technical disciplines, it was treated as a tool. These framings shaped how students interacted with the system, what kinds of questions they asked, and how they evaluated its trustworthiness.

6. Discussion: AI-Literacy, Control, and the Future of Learning

Our findings demonstrate that AI can be integrated into university education without replacing human teaching or compromising academic standards. By embedding the assistant within course structures and linking it to curated materials, we offer a model for responsible AI use that preserves institutional control and academic quality. This stands in contrast to the unregulated, unsupervised use of LLM tools, which often encourage surface learning and can undermine critical engagement. The key to successful institutional implementation lies in student motivation and self-regulation, which are crucial for effective learning in general, not just digital systems. When designed with these factors in mind, AI tools can support rather than hinder active learning.

6.1. AI as a Diagnostic Lens: Revealing Institutional Tensions

More unexpectedly, the integration of the assistant also served as a diagnostic lens, surfacing pre-existing institutional and pedagogical tensions. Teachers' individual feedback and collective discussions often revealed more about institutional hierarchies and assumptions than about the technology itself. Early in the period of growing awareness of student AI use, an assistant professor described failing a paper because it contained what he considered an implausibly "fancy" English word—gilded—which he interpreted as proof of AI authorship. He justified this based on his own linguistic background. As a non-native speaker educated at a Western university, he rarely has to consult a dictionary and therefore found the vocabulary suspicious. Beyond the anecdote, this interaction reveals entrenched hierarchies: a presumption that students' linguistic abilities must necessarily be lower than those of instructors, and that deviation signals malpractice. Drawing on my own experience as a student, this method of determining the presence of AI interaction is less than perfect. As our anthropologist colleague pointed out: during his university years, he himself often used paper dictionaries to find more sophisticated synonyms to improve his English, and that has never been seen as a sign of cheating or plagiarism. In Geertzian terms, a “thick description” of this episode reveals the instructor’s reaction as a window into institutional insecurities rather than a clever way of spotting LLM use in an essay.

6.2. Discipline-Specific Anxieties and Institutional Reflexivity

Four focus group discussions with university lecturers conducted in the spring semester of 2025 further underscored this point. Across disciplines, concerns clustered around control, accuracy, and the institution’s relevance, but were framed in discipline-specific ways. Statisticians questioned the system’s capacity to “count properly,” programmers worried about terminological precision, and social scientists feared repetitive relativisation. In each case, the technology became a site where existing institutional anxieties were articulated and negotiated, rather than the source of those anxieties. This highlights a key, and often overlooked, role that the emergence of AI systems can play in improving universities: beyond tutoring students, they reflect to institutions the tacit assumptions, inconsistencies, and power dynamics embedded in their pedagogical cultures.

6.3. Toward institutional AI literacy

Looking across these dynamics, the three orientations outlined in the introduction — non-use, misuse, and constructive use — remain a useful lens. Non-use continues to manifest through blanket bans and defensive pedagogical reflexes; misuse emerges where interaction remains unstructured and unreflective; and constructive use develops when institutions actively shape how AI is framed, deployed, and understood. Our findings suggest that moving toward constructive use requires more than technical integration: it depends on cultivating institutional AI literacy, making tacit hierarchies visible, and deliberately embedding these systems within pedagogical structures that support both trust and critical engagement. In this sense, AI is not only a pedagogical tool but also a medium through which the future of academic authority, control, and learning is being actively negotiated.

7. Conclusions

The controlled integration of generative AI into higher education provides a unique vantage point for observing how technologies are appropriated, contested, and normalized within institutional settings. By combining ethnographic inquiry with interaction data, this study has shown that AI does not simply act on educational environments—it reflects and amplifies the pedagogical, disciplinary, and institutional logics already in place. Moving forward requires institutions to go beyond technical adoption: it demands cultivating AI literacy, reflecting on tacit hierarchies, and deliberately framing pedagogy. In this sense, AI in education is both a tool for learning and a medium through which the future shape of academic authority, control, and collaboration is being worked out in practice.

References

Asch, S. E. (1952). Social psychology. Englewood Cliffs, NJ: Prentice-Hall.

Creswell, J. W., Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). Thousand Oaks, CA: Sage.

Dicks, B., Mason, B., Coffey, A., Atkinson, P. (2005). Qualitative research and hypermedia: Ethnography for the digital age. London: Sage.

Németh, R., Tátrai, A., Szabó, M., Tamási, Á. (2024). Using a RAG-enhanced large language model in a virtual teaching assistant role: Experiences from a pilot project in statistics education. Hungarian Statistical Review, 7(2), 3-27.

Németh, R., Tátrai, A., Szabó, M., Zaletnyik, P. T., Tamási, Á. (2025). Exploring the use of retrieval-augmented generation models in higher education: A pilot study on artificial intelligence-based tutoring. Social Sciences Humanities Open, 12, Article 101751

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., Tacsan, J. (2016). Digital ethnography: Principles and practice. London: Sage.

Reeves, T. C. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenney N. Nieveen (Eds.), Educational design research (pp. 52-66). London: Routledge.

Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly and ChatGPT (OpenAI, GPT-5) for text compression and grammatical editing. After using these tools, the author(s) carefully reviewed and revised the content, and take full responsibility for the final version of the manuscript.