

From Voluntary Practice Towards Actionable Learning Analytics: Architecture, Feature-Level Use, and Design Implications for an AI-Supported Learning Platform

Radka Stolariková^{1*}, Aneta Vadovičová¹, Monika Dobiášová¹, Petr Valášek¹, Jiří Votava¹

¹ Institute of Education and Counselling, Czech University of Life Sciences Prague, Prague, Czech Republic

Abstract

INTRODUCTION: Voluntary use of AI-supported learning platforms may differ across features, while aggregate logs cannot show whether teachers act on analytics.

OBJECTIVES: To evaluate AI Sensei's architecture, logged activity across platform features, self-reported feature use and student feedback.

METHODS: An observational pilot analysed logs for 205 students and an anonymous questionnaire (N = 100). Training quizzes and feedback were programmed; chat used generative AI.

RESULTS: All 205 students submitted training quizzes and self-check tests (1,999 and 481 submissions, respectively); 164 (80.0%) sent at least one chat message. In the separate anonymous questionnaire, 91% selected quizzes/tests and 17% the chat assistant; 94% agreed that combining teacher support with AI Sensei was more beneficial than teaching without the platform.

CONCLUSION: Structured practice merits refinement; teacher-facing analytics require evaluation for actionability.

Keywords: AI-supported learning, learning analytics, higher education, event logging, formative feedback, teacher-facing dashboards

Received on 03 September 2026, accepted on 05 October 2026, published on 06 October 2026

Copyright © 2026 Radka Stolariková *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/dtip.14898

1. Introduction

1.1. Voluntary practice and feedback as a learning-design problem

Digital learning environments can extend opportunities for practice beyond scheduled teaching, but voluntary use depends on students choosing to engage with optional activities alongside other study demands. This creates a

practical design challenge: learning activities must be accessible and suitable for repeated independent use.

Formative feedback is important in this context because it provides information that learners can use to adjust subsequent performance [1]. Digital practice environments can operationalise this principle through short, repeatable activities with immediate feedback, allowing students to monitor their responses and identify areas requiring further review.

Retrieval practice offers a complementary rationale. Repeated low-stakes testing can support learning and longer-

*Corresponding author. Email: stolarikova@ivp.czu.cz

term retention, particularly when practice includes feedback [2]. The effectiveness of retrieval practice also depends on implementation; for example, a comparison conducted without feedback found no consistent advantage of very short answer questions over multiple-choice questions for knowledge retention [3]. The monitoring and adjustment associated with repeated practice and feedback are consistent with core processes described in self-regulated learning models [4]. When available outside scheduled sessions, optional digital practice also allows students some discretion over when and how often they practise. Related higher-education research has associated teacher autonomy support with academic engagement among university students [5]. In the present study, however, self-regulated learning and autonomy-related mechanisms are used only as interpretive context for voluntary practice and were not directly measured.

1.2. Feature use and responsible deployment of AI-supported learning

Embedding voluntary practice in an AI-supported system introduces a further design challenge: use cannot be assumed simply because a tool is available. A recent meta-analysis in tertiary education found that AI adoption is associated with individual, contextual, and technological factors, with substantial variation across studies [6]. Across the included studies, use intentions, self-beliefs, attitudes, and social influences were among the strongest correlates, while prior AI experience moderated some associations [6]. These findings indicate that engagement with AI-supported tools is context dependent and should not be inferred from favourable attitudes or system availability alone.

Patterns of use are only one consideration in the responsible deployment of AI-supported learning systems. Responsible engagement with generative AI has been framed as requiring competencies that extend beyond basic technical use. A recent competency model of generative AI literacy encompasses technical and interaction competencies alongside critical evaluation and ethical awareness [7]. In a survey of university students in China, generative AI self-efficacy and trust were associated with acceptance, while perceived risk moderated the association between trust and acceptance [8]. These constructs were not measured in the present study and are therefore not used to explain differences in self-reported feature use.

Responsible deployment also raises governance questions. A content analysis of publications on generative AI governance in higher education found that transparency, privacy, and fairness received more attention than technical issues such as robustness [9]. A systematic review of human-centred learning analytics and AI in education similarly emphasises stakeholder agency, human control, safety, reliability, and trustworthiness [10]. These concerns are directly relevant to platforms that log student activity: in a human-centred design, automated functions should complement rather than replace pedagogical judgement,

while data collection and access should be transparent to students [10].

1.3. From interaction data towards actionable learning analytics

Recorded activity is not the same as analytical insight. A systematic review of online trace data indicates that behavioural records can reveal aspects of learning activity but interpreting them requires theoretical grounding and attention to methodological issues such as time segmentation and validity [11]. Aggregate counts are useful for describing activity volume and intensity, yet they provide limited information about the sequence or purpose of individual actions. In a recent study of LLM-supported grammar learning, higher- and lower-gain learners differed in the temporal patterns and combinations of grammar-learning strategies they used with chatbot support [12]. Although that study addressed a specific language-learning context, it illustrates why frequency alone may provide an incomplete account of how a system is used.

For instructors, analytical indicators become more useful when they can be interpreted in relation to a plausible pedagogical response. A systematic review of teacher-facing learning analytics dashboards found that many dashboards were designed to increase teachers' awareness of student activity, while offering more limited support for actionable intervention [13]. Complementary evidence from student-facing dashboard research points to a gradual move towards theory-informed designs that draw on more than activity data and connect analytics more explicitly to learning processes [14]. Together, these reviews suggest that increasing the volume of collected data is insufficient by itself; analytical value also depends on selecting indicators that can be interpreted in relation to a relevant teaching or learning decision.

Recent higher-education deployments also illustrate how behavioural traces can be combined with students' self-reports and qualitative accounts. GPTutor, for example, was evaluated using back-end interaction records, survey data, and follow-up focus groups [15]. A semester-long economics study combined assignment-level AI-use telemetry with academic outcomes and student surveys; more frequent use of AI hints was associated with more attempts but not with higher homework, exam, or final course scores [16]. The present study therefore does not claim novelty merely from combining logs and questionnaire data. Instead, this extension examines feature-level log indicators and self-reported feature use alongside the platform architecture and event-logging design to inform priorities for the next iteration of AI Sensei.

1.4. Present study and contributions

An earlier conference report [17] analysed the same two pilot datasets: platform records for 205 students and 100 valid anonymous questionnaire responses. It reported aggregate

training quiz and self-check test indicators, questionnaire evaluations, an association between reported use frequency and a mean evaluation score, and descriptive summaries of open-ended feedback. The present extension adds per-student distributions of training quiz and self-check test submissions and correct-answer totals (Table 1), together with log-derived distinct-user and interaction counts for study materials, presentations, audio resources and chat (Section 4.2). Table 1 retains the aggregate indicators reported in [17]; Table 3 revisits the B1–H1 evaluation items, and the O1 part of Table 5 revisits the perceived-benefit categories. Newly reported questionnaire results comprise the feature-use distribution (Table 2), use-frequency distribution (Table 4), category counts for O2 suggestions (Table 5), and the exploratory association between the number of U0 options selected and the overall evaluation index. Table 6 adds a literature-based comparison of AI functions and the evidence examined in two related platform evaluations. Compared with [17], the present article clarifies the questionnaire invitation flow, corrects the mean training quiz count and two item agreement percentages, and recalculates the use-frequency correlation from the same 100 valid questionnaire responses. The platform offered low-stakes training quizzes based on the study materials, immediate feedback identifying correct answers with brief explanations, self-check tests, and a chat-based AI assistant. Use remained voluntary throughout the deployment and had no bearing on formal course assessment. Selected platform interactions were recorded in event logs under pseudonymised identifiers. Because the design was observational, the findings are interpreted as descriptive and associational rather than as evidence of causal effects on learning.

The analysis drew on two separate but complementary data sources. Platform logs from 205 students with recorded platform activity provided submission counts and per-student distributions, accuracy indicators, and distinct-user and interaction counts for the other recorded platform functions. An anonymous post-exam questionnaire completed by 100 students captured self-reported use, evaluative perceptions, and open-ended feedback on perceived benefits and suggested improvements. Questionnaire responses could not be linked to pseudonymised log records at the individual level; the two datasets were therefore analysed as complementary sources of evidence rather than as matched individual-level data.

These additional analyses help specify priorities for the next iteration of AI Sensei. The study did not examine whether teachers used dashboard indicators to inform pedagogical decisions; the actionability of teacher-facing analytics therefore remains to be evaluated.

Accordingly, the study addresses the following research questions:

RQ1. What do platform logs indicate about recorded interactions across platform features, the volume and per-

student distribution of training quiz and self-check test submissions, and accuracy in these activities during the pilot deployment?

RQ2. How do questionnaire respondents report their use of AI Sensei features and frequency of use?

RQ3. What perceived benefits and suggested improvements are identified in student evaluations and open-ended feedback, and how do these findings inform priorities for the next iteration of AI Sensei?

2. AI Sensei Platform: Design, Architecture, and Data Flow

2.1. Functional scope of the pilot

AI Sensei is a modular web application that integrates learning materials, practice activities, and AI-assisted interaction within an event-logged learning environment. During the pilot, students could use study materials, presentations, audio resources, low-stakes training quizzes, self-check tests, and a chat-based AI assistant. The platform's programmed functions provided training quiz questions and immediate feedback based on the study materials; the chat assistant was its generative AI function. Teachers had access to a dashboard presenting aggregated activity and quiz/test indicators. The evaluation was restricted to functions available during the pilot semester; planned practical-training modules and other extensions were outside its scope.

2.2. System architecture

The platform uses a multi-tier architecture comprising a responsive single-page application/progressive web application (SPA/PWA) front end, serverless backend services, an AI orchestration layer, a NoSQL document database, and an analytics layer. The front end supports desktop and mobile access, while backend services process application requests and execute application logic. During the pilot, training quiz questions and immediate feedback were provided by programmed platform functions using the study materials, without calling the generative model. For chat interactions, the orchestration layer supplied student-entered messages and relevant study-material context to Google Gemini 2.5 through Vertex AI for response generation. The specific Gemini 2.5 variant used during the pilot could not be established from the available technical information. Operational data are persisted in the NoSQL database, while selected event-log records are processed through the analytics layer for aggregate analysis and reporting. The implemented platform architecture and analytical data flow are summarised in Figure 1.

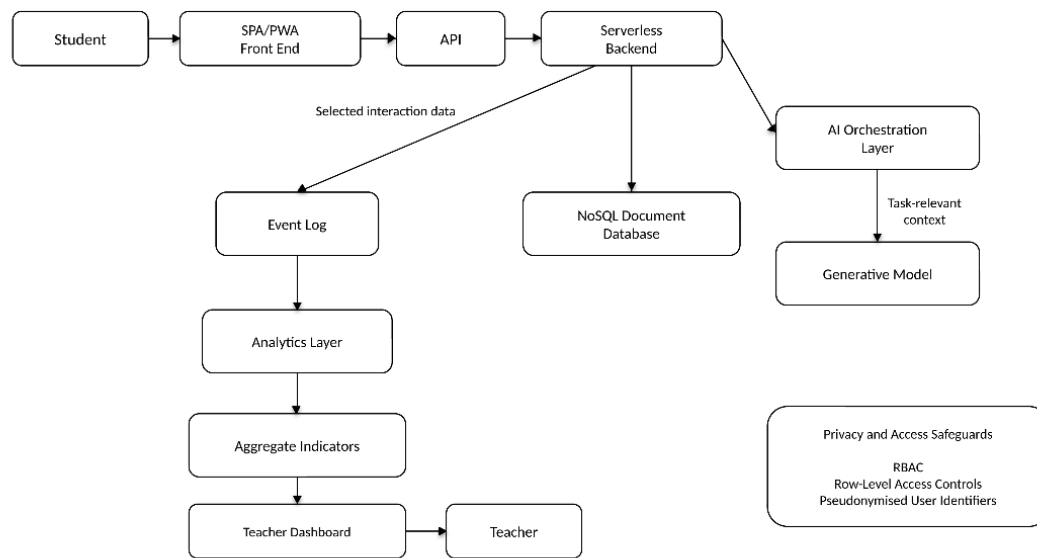


Figure 1. AI Sensei architecture and data flow: generative chat; programmed quizzes and feedback

2.3. Event logging and analytical data flow

Selected user interactions are transmitted through the application programming interface (API) to the backend for processing. Following backend processing, operational data are persisted, while selected interaction data are written to the event log. The analytics layer processes these records on demand or periodically to produce aggregate metrics for reporting and analysis.

The event-logging specification gives two examples of time-stamped records linked to pseudonymised user identifiers: `quiz_attempt_submit` (timestamp, `user_id_hash`, `quiz_id`, `score`, `duration_ms`) and `content_interaction` (timestamp, `user_id_hash`, `content_type`, `action`). The archived pilot-period records were aggregated to obtain the number of distinct students submitting training quizzes and self-check tests, the distribution of submissions per student, total valid submissions, and accuracy. For study materials, presentations, audio resources and the chat assistant, the analysis also counted distinct students and recorded interactions. These log-based indicators and the anonymous questionnaire responses were analysed separately and could not be linked at the individual level. Individual event sequences, trajectories, transitions between activities, and temporal interaction pathways were not analysed.

2.4. Privacy safeguards and access control

Analytical logs use pseudonymised user identifiers. Role-based access control (RBAC) and row-level access controls

restrict access to platform data, while research extracts used for analysis are prepared without direct personal identifiers. According to the platform's technical specification, requests to Google Gemini 2.5 through Vertex AI included relevant study-material text and student-entered chat messages, while user metadata were excluded. Under the Vertex AI terms reported by the implementation team, Google did not use customer prompts to train models without the customer's prior permission. Study-specific consent procedures and research-data handling are described in Section 3.5.

3. Materials and Methods

3.1. Study design and setting

The pilot ran from October 2025 to January 2026 at a higher-education institution in the Czech Republic. AI Sensei was used in two first-year courses within an education-related bachelor's programme across full-time and combined study modes (the latter combining scheduled in-person sessions with substantial independent study). The platform supplemented regular teaching and could be used during selected classes as well as independently outside scheduled teaching. The study was observational, with no randomisation or control group, and was designed to describe logged activity across platform functions, accuracy in training quizzes and self-check tests, self-reported feature use, and student evaluations, rather than to estimate causal effects.

3.2. Participants and deployment

All 205 students enrolled in the two pilot courses had access to AI Sensei and had at least one recorded login during the pilot. Use of AI Sensei was voluntary, carried no incentives or penalties, and had no bearing on course assessment.

An anonymous questionnaire was administered after the relevant course examination. All 205 students were invited by email, and 140 of them were additionally approached in person after the examination. Eligibility did not depend on using the platform's learning functions. Of 108 returned questionnaires, eight were excluded because they were incomplete, leaving 100 valid responses (48.8% of the students invited by email). Forty respondents were enrolled in the full-time study mode and 60 in the combined study mode. Because the questionnaire was anonymous, responses could neither be linked to individual platform log records nor attributed to the 140 students approached in person.

3.3. Data sources and measures

Platform logs. Archived pilot-period records were used to determine, separately for training quizzes and self-check tests, the number of distinct students with at least one valid submission, the total number of valid submissions, and the distribution of submissions per student. All 205 students were included in the per-student distributions and in the denominator of the mean quiz count. The records were also used to calculate distinct-user and interaction counts for study materials, presentations, audio resources and the chat assistant.

A student could submit several different quizzes or self-check tests, but could validly submit a particular quiz or test only once. Reopening an unfinished activity did not count as a new submission. Each valid submission contained 20 equally weighted questions. Its percentage score was calculated as 100 times the number of correct answers divided by 20; unanswered questions received zero points. Mean accuracy was the arithmetic mean of the percentage scores across valid submissions, calculated separately for training quizzes and self-check tests. Unsubmitted activities were excluded, and duplicate technical records were removed using submission identifiers.

For the other functions, an interaction was defined as opening a specific study material, opening a presentation, starting audio playback, or successfully sending a student chat message, respectively. Slide changes, pausing and resuming the same audio playback, and system-generated chat replies were not counted as additional interactions.

Questionnaire. For the analyses reported here, U0 captured respondents' self-reported use of platform components (study materials, presentations, quizzes/tests, audio resources, chat assistant, or other); multiple selections were allowed. Unlike the log measures, U0

combined training quizzes and self-check tests in a single response option. U1 captured self-reported frequency of use in five ordered categories: once, 2–3 times, approximately weekly, several times per week, and almost daily. U0 offered no 'none' option and U1 no 'never' option. These questionnaire responses alone therefore cannot establish the proportion of eligible students who used each learning function.

Twelve evaluative items (B1–H1) were rated from 1 (*strongly disagree*) to 5 (*strongly agree*). They covered understanding and clarity, motivation and engagement, independent learning and control over pace, feedback, the perceived value of combining teacher support with AI Sensei, support for understanding rather than answer provision, academically appropriate use, and willingness to use the platform in other courses. The evaluative items were developed for the present study and were not part of a previously validated scale. For the two exploratory correlational analyses, an overall evaluation index was calculated as the equally weighted arithmetic mean of the 12 items. It was used as a descriptive summary across several aspects of respondents' evaluations, not as a measure of a single latent construct or a validated psychometric scale.

Open-ended feedback. O1 asked respondents to identify the main benefit of AI Sensei; O2 asked what they would improve regarding its functions, clarity, content, or usability.

3.4. Data analysis

Questionnaire data were entered and checked in Microsoft Excel. Descriptive statistics and Cronbach's α were calculated in TIBCO Statistica 14.0; the two exploratory Spearman correlations were recalculated from the 100 anonymised questionnaire responses in Python using NumPy and SciPy. Platform indicators and questionnaire responses were summarised descriptively. For B1–H1, means, standard deviations, and agreement percentages were reported. Agreement was defined as ratings of 4 or 5. Individual-item summaries were the primary evaluation results; the overall index was used only for exploratory correlational analyses. Cronbach's α was calculated to describe the items' internal consistency; it was not treated as evidence that the overall evaluation index was unidimensional or validated.

Two exploratory Spearman rank correlations examined associations of the overall evaluation index with self-reported frequency of use (U1) and with the number of U0 response options selected. The U0 count did not measure the amount of activity within any selected feature. Tests were two-tailed with $\alpha = 0.05$.

Open-ended responses were categorised inductively for descriptive purposes. One author categorised the responses, and a second member of the research team checked the resulting category assignments. The check did not identify any category assignments requiring revision.

Categories were developed after reviewing the responses rather than specified *a priori*. Each substantive response was assigned to one primary category so that each respondent contributed only once to the category frequencies; additional points within the same response were not counted separately. Blank responses were recorded separately. For O2, explicit “no change” responses and other answers without a substantive improvement suggestion were distinguished from substantive suggestions; only the latter were assigned improvement categories. Category frequencies were reported descriptively.

3.5. Ethics and data protection

The study protocol and evaluation procedures, including their ethical aspects, were reviewed and approved by the authors’ institution prior to implementation. Participation in the post-exam questionnaire was voluntary. Before completing the questionnaire, respondents received information about the evaluation’s purpose, data handling, anonymity and voluntary participation, and provided informed consent. The questionnaire collected no direct personal identifiers or IP addresses. Research extracts derived from platform logs contained pseudonymised identifiers. Questionnaire and log data were handled separately and could not be linked at the individual level. Research data were stored on institutional storage in accordance with the applicable institutional policy.

4. Results

4.1. Platform activity and accuracy indicators

All 205 students with recorded platform activity submitted at least one training quiz and one self-check test (100% for each function). In total, they made 1,999 valid training quiz submissions and 481 valid self-check test submissions, corresponding to 9.8 training quizzes per student. The median number of quiz submissions per student was 10 (Q1–Q3: 8–11; range: 5–15). For self-check tests, the median was 2 (Q1–Q3: 2–3; range: 1–4): 20 students submitted one test, 104 submitted two, 71 submitted three, and 10 submitted four.

Of 39,980 quiz questions in valid submissions, 27,346 were answered correctly, yielding a mean submission accuracy of 68.4%. Of 9,620 self-check test questions in valid submissions, 7,138 were answered correctly, yielding 74.2%. Because each valid submission contained 20 equally weighted questions, these proportions equal the arithmetic means of the submission-level percentage scores (Table 1).

Table 1. Training quiz and self-check test activity and accuracy during the pilot

Indicator	Training quizzes	Self-check tests
Students with at least one valid submission, n (% of 205)	205 (100%)	205 (100%)
Valid submissions, n	1,999	481
Submissions per student, mean	9.8	2.3
Submissions per student, median (Q1–Q3)	10 (8–11)	2 (2–3)
Submissions per student, range	5–15	1–4
Correct answers / questions in valid submissions	27,346 / 39,980	7,138 / 9,620
Mean submission accuracy	68.4%	74.2%

Note: Per-student counts include all 205 students. Q1 and Q3 denote the first and third quartiles. Means are rounded to one decimal place.

4.2. Logged and self-reported use of platform features

Platform records showed that 195 of the 205 students opened study materials (95.1%; 2,340 recorded openings), 184 opened presentations (89.8%; 1,104 openings), 123 started audio playback (60.0%; 615 starts), and 164 sent at least one chat message (80.0%; 1,312 student messages). These event counts do not establish whether students studied an opened resource or completed an audio recording.

In the separate anonymous questionnaire, quizzes/tests were selected by 91 of 100 respondents, followed by study materials (59), presentations (42), the chat assistant (17), audio resources (14), and “Other” (2) (Table 2). Log-derived activity and self-reported feature use differed markedly. For example, 164 of 205 students sent a chat message according to the logs, whereas 17 of 100 questionnaire respondents selected the chat assistant. Similarly, 91 respondents selected the combined quizzes/tests option although all 205 students had submitted both types of activity in the logs. The questionnaire and log records could not be linked individually, so the reasons for these differences could not be determined.

Table 2. Self-reported use of AI Sensei platform components (N = 100)

Platform component	n	%
Quizzes/tests	91	91
Study materials	59	59
Presentations	42	42
Chat assistant	17	17

Audio resources	14	14
Other	2	2

Note: Respondents could select more than one component.

4.3. Student evaluations of AI Sensei

Internal consistency of the 12 evaluative items (B1–H1) was $\alpha = 0.894$. Mean ratings ranged from 3.76 to 4.35. Agreement was highest for the perceived value of combining teacher support with AI Sensei (F1; $M = 4.35$, $SD = 0.63$; 94%), followed by clarity of explanations (B2; $M = 4.27$, $SD = 0.72$; 89%). Agreement reached 86% for both support for understanding rather than ready-made answers (G1; $M = 4.09$, $SD = 0.83$) and willingness to use AI Sensei in other courses (H1; $M = 4.32$, $SD = 0.84$). The lowest agreement was recorded for increased motivation to study (C1; $M = 3.76$, $SD = 1.06$; 62%) (Table 3).

Table 3. Student evaluations of AI Sensei (B1–H1; $N = 100$)

Item	Mean	SD	Agreement (%)
B1: Helped to understand taught content	4.11	0.95	83
B2: Explanations were clear and well structured	4.27	0.72	89
C1: Increased motivation to study	3.76	1.06	62
C2: Supported more active engagement in learning	3.93	0.96	74
D1: Supported independent learning	3.95	0.85	76
D2: Improved control over learning pace	3.92	0.84	75
E1: Feedback was quick and useful	3.99	0.85	75
E2: Feedback clarified what needed correction and why	4.18	0.73	83
F1: Combining teacher support with AI Sensei was more beneficial than teaching without AI Sensei	4.35	0.63	94
G1: Supported understanding rather than providing ready-made answers	4.09	0.83	86
G2: Supported fair learning without encouraging cheating	4.16	0.84	81
H1: Willingness to use AI Sensei in other courses	4.32	0.84	86

Note: Agreement represents ratings of 4 or 5 on the five-point response scale.

4.4. Self-reported frequency of use and associations with the overall evaluation index

The most common self-reported frequency of AI Sensei use was approximately weekly (44%). A further 21% of respondents reported using the platform several times per week and 7% almost daily. Twenty-six per cent reported using it two or three times during the course, while 2% reported using it only once (Table 4).

Table 4. Self-reported frequency of AI Sensei use (U1; $N = 100$)

Frequency of use	n	%
Once	2	2
2–3 times	26	26
Approximately weekly	44	44
Several times per week	21	21
Almost daily	7	7

In exploratory analyses, self-reported frequency of use (U1) was positively associated with the overall evaluation index (Spearman's $r_s = 0.379$, $p < 0.001$). The number of U0 response options selected was also positively associated with the index ($r_s = 0.361$, $p < 0.001$).

4.5. Open-ended feedback and suggested improvements

Of the 100 respondents, 37 left O1 blank. Among the 63 substantive responses to O1, the most frequently identified perceived benefits were practice and repetition through quizzes ($n = 20$) and clarity and structure ($n = 17$). Learning pace and control was the next most frequent category ($n = 7$), while the remaining categories were identified less often (Table 5).

For O2, 45 responses were blank, six explicitly indicated no suggested change, and nine others did not contain a specific improvement suggestion. Among the 40 substantive suggestions recorded for O2, the most frequent suggestions concerned additional quizzes and greater item variety ($n = 11$) and more study materials or broader exam coverage ($n = 10$). Other recurring suggestions concerned interface clarity and navigation ($n = 5$), audio and podcast improvements ($n = 4$), and platform availability or access ($n = 3$); the remaining categories were reported by one or two respondents (Table 5).

Table 5. Open-ended feedback on perceived benefits and suggested improvements

Perceived benefits (O1; n = 63)	n	Suggested improvements (O2; n = 40)	n
Practice and repetition through quizzes	20	More quizzes/tests and greater item variety	11
Clarity and structure	17	More study materials and broader exam coverage	10
Learning pace and control	7	Interface clarity and navigation	5
Feedback and guidance	4	Audio and podcast improvements	4
Learning resources and formats	3	Availability and access	3
Anytime access	2	Progress tracking and results visibility	2
Confidence and reduced stress	2	Scheduling and calendar	2
Exam preparation	2	Language and tone	1
Motivation and engagement	2	Mobile and technical issues	1
Other	2	Notes and communication interface	1
Self-study and autonomy	1		
Self-testing	1		

Note: Each substantive response was assigned to one primary category. Frequencies therefore sum to n = 63 for O1 and n = 40 for O2. For O2, the categorical summary includes only substantive suggestions; blank responses and replies without a specific suggestion were excluded.

5. Discussion

5.1. Principal findings

This observational pilot examined AI Sensei as a voluntary supplementary learning platform used within routine higher-education teaching. The logs and questionnaire responses yielded different feature-use profiles, particularly for the chat assistant; among respondents, quizzes/tests were selected more often than the chat assistant or audio resources. Evaluations were generally favourable, particularly for combining teacher support with AI Sensei and for the clarity of explanations. In exploratory analyses confined to the questionnaire, reported use frequency and the number of U0 response options selected were each positively associated with the overall evaluation index. These associations do not establish their direction or a causal relationship between use and evaluation.

The most common perceived benefits in the open-ended responses were practice and repetition, followed by clarity and structure; suggestions focused on more varied practice

activities and broader study materials. Together with self-reported feature use, these responses inform priorities for the next iteration of AI Sensei. More broadly, AI use and its relationship to student evaluation or learning should be interpreted within its pedagogical context rather than inferred from system availability or overall use alone [6,16]. The event-logging and dashboard architecture enabled aggregate reporting during the pilot; whether teachers used those indicators to inform pedagogical decisions was not evaluated.

5.2. Voluntary practice, formative support, and self-regulated learning

The high self-reported use of quizzes/tests, together with the prominence of practice and repetition in the open-ended responses, identifies structured practice as a prominent aspect of respondents' reported use of AI Sensei. This pattern is consistent with the platform's formative design: use was voluntary, platform activities did not contribute to formal course assessment, and training quizzes provided questions based on the study materials with immediate feedback identifying correct answers and providing brief explanations. Research on formative feedback similarly emphasises the value of feedback that provides learners with information they can use to improve subsequent performance [1]. The opportunity to practise, receive feedback, and adjust subsequent learning activity is also compatible with core processes described in self-regulated learning models [4].

The prominence of quizzes and practice-related feedback is also consistent with retrieval-practice research. Repeated retrieval can function as a learning activity in its own right, particularly when accompanied by feedback [2]. At the same time, the present results do not show that the particular quiz format used in AI Sensei was superior to other forms of retrieval practice. In a small within-student comparison conducted without feedback, very short answer and multiple-choice practice questions did not differ significantly in later retention [3]. The high self-reported use of quizzes and requests for additional items therefore support the interpretation that structured practice was valued by respondents, but they do not provide evidence of improved learning outcomes.

Students also retained discretion over whether, when, and how often to engage with the platform. Respondents' positive evaluations of support for independent learning and control over learning pace are consistent with conditions that can support self-regulated learning. However, self-regulation was not measured as an outcome, and aggregate activity indicators cannot establish the cognitive or metacognitive processes underlying observed behaviour. Systematic evidence on online trace analysis likewise emphasises that behavioural traces require explicit theoretical linkage and attention to timing and context if they are to be interpreted as indicators of self-regulated learning [11]. The present findings therefore support interpreting AI Sensei as providing opportunities for self-

regulated practice, rather than as demonstrating an effect on self-regulated learning.

5.3. Feature-level patterns of use

Platform logs and anonymous questionnaire responses gave different profiles of feature use. For example, 164 of 205 students sent at least one chat message, whereas only 17 of 100 questionnaire respondents selected the chat assistant; all 205 students submitted both a training quiz and a self-check test, whereas 91 respondents selected the combined quizzes/tests option. These differences cannot be explained by the different denominators alone. Because questionnaire responses could not be linked to log records, differences between the two sources could not be examined at the individual level. The questionnaire percentages therefore describe respondents’ selections, not logged use across the cohort.

The chat assistant was the platform’s generative AI function; quiz questions and feedback came from programmed functions. The combined quizzes/tests response cannot be interpreted as use of the generative model. Conversely, a recorded chat message establishes that an interaction occurred but does not establish its educational value. Factors associated with educational AI use vary across contexts [6], but the present data do not identify the reasons for the discrepancy between sources or support conclusions about usability, trust, or educational value. Studies combining interaction records with survey or outcome data also illustrate the need to interpret frequency alongside other evidence [15,16]. Future evaluations should align questionnaire options with logged event definitions and examine feature-level sequences over time.

Table 6 situates AI Sensei alongside two higher-education systems, distinguishing the role of generative AI and the evidence examined in their evaluations.

Table 6. Selected higher-education systems and evidence reported in their evaluations

Study	Generative AI function	Evidence analysed; teacher decisions
AI Sensei (present study)	Chat assistant; quizzes and feedback used programmed functions	Quiz/test submission distributions and accuracy; feature-level user and interaction counts; separate anonymous questionnaire; teacher decisions not evaluated
GPTutor [15]	Course-grounded question	Chat and exercise records, engagement

	answering, exercise and flashcard generation	questionnaire and focus groups; teacher decisions not evaluated
Macmillan Achieve AI-Tutor [16]	AI hints during economics homework	Assignment telemetry, surveys and course outcomes; teacher decisions not evaluated

Note: “Not evaluated” describes the outcomes examined in the cited studies, not the full technical capabilities of either platform.

5.4. Teacher support, platform use, and actionable analytics

The highest agreement among the evaluation items concerned the perceived value of combining teacher support with AI Sensei. This preference for combining teacher support with the platform is compatible with human-centred approaches that retain pedagogical judgement and responsibility with teachers [10]. In the pilot, AI Sensei functioned as a supplementary learning environment, while formal instruction and course assessment remained under teacher control. The result therefore indicates that questionnaire respondents valued the combination of teacher support and platform-based practice and feedback, although it does not show that this combination produced superior learning outcomes.

The role of teachers also matters for the analytical layer of the system. Event logging and the teacher dashboard provide a technical basis for teacher-facing learning analytics, but the availability of interaction data alone does not make those analytics actionable. Review evidence on teacher-facing learning analytics dashboards indicates that many designs prioritise awareness of learner activity, while providing more limited support for actionable intervention [13]. An empirical study of five university instructors showed that interpreting dashboard data involved identifying patterns, explaining their significance in the course context, and deciding whether and how to respond [18]. A design-based evaluation of a teacher-facing dashboard in university courses found that involving teachers in its development and aligning indicators with pedagogical needs could support its use in practice [19]. Research on student-facing dashboards also points towards theory-informed indicators linked to learning processes [14]. The key design challenge is therefore not to maximise the number of indicators, but to connect interpretable signals with pedagogical questions and decisions that matter to teachers.

The pilot documented logged activity by platform function and quiz/test accuracy; it did not examine teacher interpretation or instructional decisions. A prospective pathway would first validate feature-level and temporal indicators, then examine how teachers interpret them and whether they use them to make specific instructional

decisions. For example, a reliable future indicator of repeated difficulty with particular material could prompt a revised explanation or additional practice; whether that decision improves learning would require subsequent evaluation. Sequence analysis and any predictive or risk indicators would require separate validation and appropriate data governance before use.

5.5. Design implications for the next iteration

Taken together, platform logs, self-reported feature use and open-ended feedback provide an empirical basis for identifying priorities for the next development cycle. The strongest evidence concerns further development of the structured practice layer. Quizzes/tests were the most widely reported component, practice and repetition formed the most frequently reported benefit category, and requests for more quizzes/tests and greater item variety were the most common suggestions for improvement. These converging findings support expanding the diversity and coverage of structured practice activities as a near-term development priority. A second development area concerns learning resources and usability, reflecting requests for broader study materials, clearer navigation and access, better visibility of progress and results, and improvements to audio and podcast resources. The marked differences between logged and self-reported use of the chat assistant and audio resources warrant closer examination of these functions before decisions about substantial expansion or removal.

The separation of user-facing, storage, event-logging, and analytics functions provides a technical basis for iterative refinement of platform components and analytical indicators, while access controls restrict access to platform data. Extending the analytical layer, however, also increases the importance of data minimisation, access control, and careful interpretation of the resulting indicators. Higher-education AI governance literature identifies privacy, transparency, and oversight as recurring considerations in responsible institutional deployment [9], while human-centred learning analytics emphasises human agency, safety, reliability, and trustworthiness in data-intensive educational systems [10]. The platform uses pseudonymised log identifiers and role-based and row-level access controls; its technical specification also describes restricting model requests to task-relevant text. The technical specification identifies study-material text and student-entered text as input categories and excludes user metadata. It does not establish the exact contents of individual pilot requests or whether and for how long those requests were retained by Google.

The modular architecture may support deployment in other settings, but the pilot documents operation with 205 students with recorded platform activity at one institution. It did not assess load or system performance, deployment costs, portability to other institutions, or integration with a learning management system (LMS). Wider deployment

would require technical testing and assessment of whether additional analytics are interpretable and useful to teachers. AI Sensei is also being trialled in a secondary-school setting, but that deployment is outside the present analysis. Whether its event logging and teacher-facing analytics can inform instructional decisions in other educational or workplace training contexts requires separate evaluation.

5.6. Limitations and future research

The findings should be interpreted within the limits of an observational pilot conducted in two first-year courses within one bachelor's programme at a single institution. Platform use was voluntary, there was no control group, and the study did not include pre-post measurement of learning outcomes. Consequently, the study design does not permit causal inference regarding effects of AI Sensei on learning outcomes, engagement, or self-regulated learning, and the transferability of the observed patterns of use remains to be examined across other courses, disciplines, and institutions. Because platform use and questionnaire participation were voluntary (100 valid responses from 205 students invited by email), self-selection and response bias cannot be excluded. Although the 12 study-specific evaluative items showed internal consistency, they covered several aspects of students' experiences; their arithmetic mean was therefore treated as a descriptive overall evaluation index in exploratory analyses, not as a validated or demonstrably unidimensional scale.

A second limitation concerns the relationship between platform-log and questionnaire data. Because the questionnaire was anonymous, it could not be linked at the individual level to pseudonymised platform-log records. U0 feature use and U1 frequency of use were self-reported, whereas the log analyses used archived records to derive submission distributions and feature-level interaction counts. U0 offered no 'none' option and U1 no 'never' option; moreover, respondents' feature-use selections differed markedly from logged activity. The anonymous design prevented assessment of these differences for individual students. Because U0, U1, and the overall evaluation index were derived from the same post-exam questionnaire, the observed associations may also be affected by recall and common-method bias. Future studies could use privacy-preserving linkage, where ethically and technically appropriate, and analyse feature-level event sequences over time. Such designs would enable more detailed analysis of how students move between platform components and how patterns of use develop over time, while addressing methodological concerns about the interpretation and temporal organisation of online traces [11].

The open-ended responses were brief and were analysed using descriptive categorisation; they were not intended to provide an in-depth account of user experience. Assigning each substantive response to one primary category meant that additional issues mentioned in the same answer were

not counted separately. Interviews or focus groups could help explain why respondents' accounts of feature use differed from logged activity. The pilot did not evaluate how teachers used the dashboard in practice.

6. Conclusion

This observational pilot examined platform-log indicators alongside anonymous questionnaire responses from a voluntary deployment of AI Sensei. The two sources yielded different feature-use profiles; respondents particularly valued combining teacher support with the platform.

Student feedback identifies more varied practice activities and broader study materials as priorities for the next iteration, while requests concerning navigation and audio quality suggest specific refinements. The chat assistant and audio resources warrant targeted evaluation before substantial expansion.

Relative to the conference report [17], this article adds per-student quiz/test submission distributions, distinct-user and interaction counts for individual features, self-reported feature-use and frequency distributions, and category counts for suggested improvements. It also describes the architecture and event-logging design in greater technical detail. The architecture provides a technical basis for further development of teacher-facing analytics, but the present study did not establish whether such indicators are actionable in pedagogical practice. Future studies should test whether validated feature-level and temporal indicators inform teachers' decisions and whether any resulting changes improve learning, including in other educational contexts.

Acknowledgements

The authors thank the students who participated in the evaluation questionnaire for their time and valuable feedback on AI Sensei.

Generative AI tools were used solely to support language editing, stylistic refinement, and improvements in clarity and readability. All scientific content and final wording were reviewed and approved by the authors, who take full responsibility for the manuscript.

References

- [1] Shute VJ. Focus on formative feedback. *Rev Educ Res*. 2008;78(1):153–189. <https://doi.org/10.3102/0034654307313795>
- [2] Brame CJ, Biel R. Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *CBE Life Sci Educ*. 2015;14(2):es4. <https://doi.org/10.1187/cbe.14-11-0208>
- [3] van Wijk EV, de Jonge M, van Blankenstein FM, Janse RJ, Langers AMJ. The battle of question formats: a comparative study of retrieval practice using very short answer questions and multiple choice questions. *BMC Med Educ*. 2024;24(1):1547. <https://doi.org/10.1186/s12909-024-06538-0>
- [4] Panadero E. A review of self-regulated learning: Six models and four directions for research. *Front Psychol*. 2017;8:422. <https://doi.org/10.3389/fpsyg.2017.00422>
- [5] Li Y, Sueb R, Said Hashim K. The relationship between parental autonomy support, teacher autonomy support, peer support, and university students' academic engagement: the mediating roles of basic psychological needs and autonomous motivation. *Front Psychol*. 2025;16:1503473. <https://doi.org/10.3389/fpsyg.2025.1503473>
- [6] Teo T, Scherer R, Fung ASK, Fung CSL. Factors associated with students' adoption of artificial intelligence technology in tertiary education: A meta-analytic review. *Educ Res Rev*. 2026;52:100804. <https://doi.org/10.1016/j.edurev.2026.100804>
- [7] Annappureddy R, Fornaroli A, Gatica-Perez D. Generative AI literacy: Twelve defining competencies. *Digit Gov Res Pract*. 2025;6(1):Article 13, 1–21. <https://doi.org/10.1145/3685680>
- [8] Zhang G, Yu T. Association between Generative AI self-efficacy and Generative AI acceptance: The mediating role of Generative AI trust and the moderating role of Generative AI risk perception. *Acta Psychol (Amst)*. 2025;261:105791. <https://doi.org/10.1016/j.actpsy.2025.105791>
- [9] Tong A, Zainol Z, Chong TS, Renganathan K. AI governance on young consumers in higher education: a content analysis of policies for generative AI. *Young Consum*. 2025;26(5):865–885. <https://doi.org/10.1108/YC-10-2024-2303>
- [10] Alfredo R, Echeverria V, Jin Y, Yan L, Swiecki Z, Gašević D, Martinez-Maldonado R. Human-centred learning analytics and AI in education: A systematic literature review. *Comput Educ Artif Intell*. 2024;6:100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- [11] Du J, Hew KF, Liu L. What can online traces tell us about students' self-regulated learning? A systematic review of online trace data analysis. *Comput Educ*. 2023;201:104828. <https://doi.org/10.1016/j.compedu.2023.104828>
- [12] Hao C, Zhang F. Understanding self-regulated grammar learning with LLM chatbot support: An epistemic network analysis of grammar learning strategy patterns. *System*. 2026;136:103879. <https://doi.org/10.1016/j.system.2025.103879>
- [13] Kaliisa R, Jivet I, Prinsloo P. A checklist to guide the planning, designing, implementation, and evaluation of learning analytics dashboards. *Int J Educ Technol High Educ*. 2023;20:28. <https://doi.org/10.1186/s41239-023-00394-6>
- [14] Paulsen L, Lindsay E. Learning analytics dashboards are increasingly becoming about learning and not just analytics - A systematic review. *Educ Inf Technol*. 2024;29(11):14279–14308. <https://doi.org/10.1007/s10639-023-12401-4>
- [15] Bai H, Lui WC, Khiatani PV. Promoting student engagement with GPTutor: An intelligent tutoring system powered by generative AI. *Int J Educ Technol High Educ*. 2025;22:77. <https://doi.org/10.1186/s41239-025-00571-9>
- [16] Mask J. Does AI help economics students learn or just finish? A classroom field study. *Int Rev Econ Educ*. 2026;52:100350. <https://doi.org/10.1016/j.iree.2026.100350>
- [17] Stolariková R, Vadovičová A, Dobiášová M, Valášek P, Votava J. AI Sensei: Design and pilot evaluation of an event-logged learning platform with analytics for higher

- education. In: Machado J, Trojanowska J, Antosz K, Leão CP, Knapcikova L, Sover A, editors. *Innovations in Industrial Engineering V. Lecture Notes in Mechanical Engineering*. Cham: Springer Nature; 2026. In press.
- [18] Wise AF, Jung Y. Teaching with analytics: Towards a situated model of instructional decision-making. *J Learn Anal*. 2019;6(2):53–69.
<https://doi.org/10.18608/jla.2019.62.4>
- [19] Kaliisa R, Dolonen JA. CADA: A teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions. *Technol Knowl Learn*. 2023;28:937–958.
<https://doi.org/10.1007/s10758-022-09598-7>