# A Multi-Criteria Software Quality Evaluation of AI Meeting Assistants for English-Medium University Lectures Using ISO/IEC 25010 and TOPSIS

Alex Pak Ki Kwok[1,*], Yao Hing Wong[2], Kwong-Cheong Wong[1], and Chee Hon Chan[1]

[1]Data Science and Policy Studies Programme, School of Governance and Policy Science, Faculty of Social Science, The Chinese University of Hong Kong, Hong Kong, China
[2]English Language Teaching Unit, The Chinese University of Hong Kong, Hong Kong, China
Email: alexkwok@cuhk.edu.hk (A.P.K.K.); yaohingwong@cuhk.edu.hk (Y.H.W.); kwongcheongwong@cuhk.edu.hk (K.-C.W.); cheehonchan@cuhk.edu.hk (C.H.C.)
*Corresponding author
Manuscript received May 28, 2025; revised June 20, 2025; accepted July 28, 2025; January 9, 2026

***Abstract*—This study aims to systematically evaluate the effectiveness of ten Artificial Intelligence Meeting Assistants (AIMAs) in supporting English-medium university lectures. The research was carried out at The Chinese University of Hong Kong and involved both teachers and students as stakeholders. Using a within-subjects design, twenty participants (twelve students and eight teaching staff, recruited through snowball sampling across multiple faculties) tested each AIMA in simulated lecture contexts. Data were collected through structured questionnaires based on the ISO/IEC 25010 software quality framework, covering nine criteria including functional suitability, performance efficiency, compatibility, usability, reliability, security, satisfaction, sustainability, and scalability. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) was used to combine participants' importance ratings and performance scores, resulting in a final ranking of the AIMAs. TOPSIS analysis of participant evaluations ranked Tl;dv most favorably, followed by Grain and Microsoft Teams. Notably, teachers rated security ($p < 0.001$) and performance efficiency ($p = 0.009$) significantly higher than students, highlighting differing user priorities. This study provides empirical benchmarks and a replicable framework for selecting educational technologies. The findings may help institutions make evidence-based decisions about using AIMAs to improve student understanding and participation in linguistically diverse classrooms.**

***Keywords*—Artificial Intelligence (AI) meeting assistants, educational technology, English-medium instruction, ISO/IEC 25010, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)**

## I. INTRODUCTION

English-Medium Instruction (EMI), which is the use of English to teach academic subjects in regions where English is not the dominant language [1], has become increasingly popular across universities worldwide as a result of the growing internationalization of higher education [2]. Data from Dearden [1] underscore this momentum: among 55 nations surveyed, two out of five reported adopting EMI policies, while nearly half had publicized official statements supporting the approach. In Central and Eastern Europe, student participation in EMI programs climbed from 9% to 25% [3]. In China, 127 universities across 25 provinces are found to have run/having on offer a total of 620 EMI-focused undergraduate programs [4]. Consequently, many students are now expected to learn in a language other than their mother tongue as educational institutions embrace greater linguistic diversity. This practice poses new challenges for students' engagement and comprehension in lecture-based settings. Students with limited proficiency in the language of instruction often face difficulties in understanding lecture content and staying engaged, which can undermine their academic performance and participation in class [5]. These challenges highlight the urgent need for creative, technologically advanced solutions that promote inclusivity and facilitate efficient learning in academic contexts with multiple languages.

Artificial Intelligence Meeting Assistants (AIMAs) appear to be a promising solution for supporting students in multilingual educational contexts, including EMI environments. These AI-driven tools can transcribe lectures, identify key topics, and generate structured summaries, thereby facilitating student comprehension in EMI [6, 7]. In particular, during lectures, AIMAs can generate live transcriptions and captions of spoken content. This real-time assistance can reduce the cognitive burden of simultaneous listening, comprehension, and note-taking that multilingual students often face. By alleviating the pressure of manual notetaking, students can focus more effectively on understanding complex concepts while simultaneously acquiring academic vocabulary in English [6]. After lectures, AIMAs can also generate structured summaries that allow students to review and deepen their understanding of both the subject matter and related English-language academic discourse. These post-lecture resources provide additional opportunities for comprehension and language development outside the classroom. Additionally, as a side benefit, AIMAs continuously record students' active participation and discussion contributions during lectures. This information can provide teachers with valuable data to assess classroom engagement and monitor the effectiveness of student interaction in EMI environments.

The growing adoption of AI in education underscores the relevance of these tools. Recent data reveal a rise in AI use in education. A Digital Education Council survey found that 86% of students worldwide use AI tools in their studies [8]. Faculty adoption is also accelerating, as 93% of higher education staff anticipate increased reliance on AI in teaching and administrative work [9]. Although AIMAs have gained ground in business, reaching a market size of USD 1.95 billion in 2023 and projected to grow to nearly USD 12 billion by 2031 [10], their effectiveness in supporting students' learning in EMI context remains under-examined.

doi: 10.18178/ijiet.2026.16.1.2479

The proliferation of commercially available AIMAs, each claiming diverse functionalities, presents a significant challenge for educators and administrators seeking solutions that effectively serve both instructional and student needs. This study attempts to address this pressing need by offering a comprehensive, stakeholder-informed, and replicable evaluation framework for AIMAs so as to provide institutions with timely, actionable insights to guide evidence-based selection and implementation. Drawing upon the ISO/IEC 25010 software quality framework [11] and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method [12], this research assesses ten prominent AIMAs based on nine comprehensive software quality and user-centric criteria. The study also explores how teachers and students prioritize these evaluation criteria when adopting AIMAs in classrooms. The primary aim is to provide a systematic and timely comparison of available AIMAs to inform current institutional decisions.

While prior studies have examined individual AIMAs or qualitatively compared a limited set of tools, most research focuses on technical attributes or single-stakeholder perspectives. They also often focused on the use of AIMAs in business contexts rather than education (see Table 1). This study is novel in a number of ways: First, it systematically evaluates ten widely used AIMAs specifically within the context of English-medium university lectures. Second, it employs the internationally recognized ISO/IEC 25010 software quality framework in conjunction with the multi-criteria decision-making technique TOPSIS to ensure a rigorous and holistic assessment. Last but not least, it incorporates the weighted priorities that both teachers and students assign to nine evaluation criteria, as the successful integration of educational technology relies on the acceptance and positive attitudes of both groups [13].

There are three main research questions guiding this study:

1) How do current AIMAs perform when evaluated against ISO/IEC 25010 criteria in supporting English-medium university lectures?
2) Which AIMAs currently demonstrate the highest overall performance when evaluation criteria are weighted according to stakeholder priorities?
3) In what ways do teachers' and students' priorities differ regarding the features and capabilities of AIMAs?

The answers to these questions could support educational institutions and instructors in making evidence-based decisions concerning the adoption of AIMAs to enhance student comprehension and engagement in linguistically diverse classrooms.

## II. Literature Review

### A. Challenges in English Medium Instruction

English-medium instruction (EMI) has expanded rapidly as universities seek to internationalize and produce graduates able to compete globally [2]. This shift, however, has introduced a new classroom challenge. The expectation that students already command the necessary English skills is widespread yet not always justified in practice. Research in Chinese universities has documented frequent mismatches between institutional language policies and the lived realities of students [4]. As such, instructors often face classrooms where language proficiency varies widely. Echoing these concerns, Li and Pei [14] provide quantitative evidence that inadequate academic English skills, especially in writing, are a major predictor of student failure in EMI settings. Their regression analysis reinforces the centrality of language readiness to students' academic success. Mai *et al.* [15] also identify students' insufficient proficiency as one of the main barriers to participation in English-speaking classes.

These challenges underscore a broader need for technological interventions that can support real-time, inclusive comprehension. However, limited empirical research explores the use of AI tools, particularly AIMAs, to address the needs of linguistically diverse students. This study seeks to fill this void by evaluating AIMAs as scalable, inclusive tools that enhance accessibility and comprehension in EMI lectures.

### B. AI Transcription and Captioning in Education

AI technologies are increasingly integrated into educational practice because of their potential to enhance personalized learning, administrative efficiency, and accessibility [16–18]. For students in EMI courses, Automatic Speech Recognition (ASR) systems, which serve as the foundational technology behind AIMAs, offer more than mere convenience. Real-time captions and transcripts can help bridge comprehension gaps and support learners who may struggle to follow the pace of instruction. Malakul and Park [7] found that AI-generated subtitles helped students manage cognitive load and improved their grasp of course content. However, these benefits are tempered by concerns about accuracy and reliability. Kuhn *et al.* [19] noted high word error rates in commercial ASR tools, particularly with specialized academic terminology, limiting their suitability for contexts requiring linguistic precision.

This gap highlights the need for robust, inclusive AI transcription solutions in higher education. To address this, this study evaluates the quality and pedagogical value of AIMAs through a rigorous empirical framework, contributing to the limited literature on AI transcription tools in authentic academic settings.

### C. AIMAs in Higher Education

AIMA were initially designed to enhance business productivity by offering features such as transcriptions and automated summaries. These functions could make meetings more efficient by enhancing communication and information retention. As educational settings become more linguistically diverse, there is increasing interest in examining how AIMAs might bridge comprehension gaps and foster student engagement, particularly in EMI contexts [6]. Scholars have investigated AIMAs across a range of instructional and professional contexts, as summarized in Table 1. For example, Cabrero-Daniel *et al.* [20] examined a custom AIMA in software development teams. They highlighted the importance of user experience and expert oversight for adoption. Additionally, Khoo *et al.* [21] compared several transcription tools using a short audio sample from YouTube. In higher education, Kwok *et al.* [6] conducted a controlled experiment with university students and observed that the class with Otter.ai perceived lower instructional clarity but expected higher grades. Despite these studies, most existing research focuses on individual tools and general meetings, often with single-stakeholder perspectives and a limited comparative scope. Multi-stakeholder, comprehensive assessments of AIMAs in educational contexts remain scarce. This gap highlights the need for systematic evaluation frameworks that integrate the perspectives of students and teachers to assess the quality and suitability of AIMAs in EMI higher education.

Table 1. Summary of recent empirical evaluations of AIMA tools

| Study | Tool(s) | Context / Domain | Methodology | Sample | Key Findings |
|---|---|---|---|---|---|
| Asthana *et al*. [35] | Custom LLM recap system | Workplace meetings | Field trial + user study | 7 knowledge workers | Recaps improved efficiency but improved contextual personalization is needed. |
| Cabrero-Daniel *et al*. [20] | Custom GPT-4 assistants | Software development | Action research | 3 Scrum teams in two software development meetings | AI assistants are capable of generating accurate and contextualized insights, exceeding some participants' expectations. Customization of AIMAs to align with both individual practitioner and team preferences is crucial. |
| Haliburton *et al*. [36] | "Walking Talking Stick" device | Outdoor meetings | Between-subjects experiment | 60 participants | The tangible device boosted task focus. The highlighting button improved turn-taking and note quality. |
| Herdiyanti [37] | Otter.ai, Qualtrics, Zoom | Research interviews | Qualitative reflective analysis | 9 transcripts per service | Non-native speech was transcribed less accurately, raising concerns about bias and accountability. Using two transcription services in parallel ensured transcripts were not lost if one failed. |
| Khoo *et al*. [21] | Otter.ai, Transcribe, TurboScribe, Whisper | General transcription | Comparative feature analysis | Evaluation on 3-min audio clip | Whisper and TurboScribe outperformed Otter.ai in accuracy; Whisper's local processing enhanced data privacy. |
| Kwok *et al*. [6] | Otter.ai | Higher education | Controlled experiment | 39 students in two classes | Otter.ai users reported reduced perceptions of instructor clarity but better expected grades; tool offered academic compensation at the cost of social dynamics. |
| Son *et al*. [38] | Custom AI transcription platform | Remote meetings | Controlled experiment | 71 global participants | Real-time transcription aided recall and focus after distractions, supporting engagement during multitasking. |

### *D. Theoretical Foundation for the Multi-Criteria Software Quality Evaluation*

When evaluating educational tools like AIMAs, it is crucial to comprehend how users accept and assess new technologies. One of the most important frameworks in this field is Davis's Technology Acceptance Model (TAM) [22]. TAM suggests that users' acceptance of a new technology is influenced by two main factors: perceived usefulness and perceived ease of use, i.e., users' expectations of performance benefits and the effort required to use the system. TAM has been expanded over time to include more context-specific factors such as compatibility, security, and satisfaction. These extensions have allowed it to explain users' technology acceptance behavior in a variety of fields, such as healthcare and education. For instance, Al-Adwan *et al*. [23] used TAM to investigate students' intentions in using metaverse-based learning platforms, while Jasrotia *et al*. [24] applied TAM to investigate user engagement in fashion e-commerce. Xue *et al*. [25] further emphasized TAM's central role in research on technology adoption within the higher education community.

While TAM offers useful insights in predicting users' adoption behavior, it does not assess the technical quality of software systems themselves. For this purpose, international software quality standards such as ISO/IEC 25010 are essential. ISO/IEC 25010 builds upon earlier quality models such as the ISO/IEC 9126 standard, which provides a set of characteristics that defines software quality. ISO/IEC 25010 includes characteristics such as functional suitability, performance efficiency, reliability, usability, security, and compatibility, which are both commonly used in software quality evaluation literature [26] and critical for educational technology in classroom settings. In educational research, Marroquin and Rodriguez [27] applied ISO/IEC 25010 to assess the quality of cloud-based e-learning platforms, highlighting its applicability in academic environments.

The nine evaluation criteria used in this study are based on the ISO/IEC 25010 standard, a widely accepted software quality model developed by international experts in systems and software engineering. To adapt these abstract categories into measurable indicators, we employed the AdEQUATE Software Quality Evaluation Model proposed by Alves *et al*. [11], which operationalizes ISO/IEC 25010 dimensions by providing validated survey instruments. Previous studies have successfully used this combined framework in a variety of fields, including higher education and telemedicine. For example, Fadhel *et al*. [28] employed the AdEQUATE model to develop questionnaire items for evaluating the quality of web-based systems in higher education settings.

However, ranking software across several quality dimensions frequently entails comparing multiple conflicting criteria, such as compatibility versus performance or usability versus security. This complexity makes traditional single-metric evaluations inadequate. Multi-Criteria Decision-Making (MCDM) techniques offer a systematic way to handle such trade-offs. Among them, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is widely recognized for its mathematical rigor, conceptual simplicity, and practical applicability [29]. TOPSIS ranks alternatives by calculating their geometric closeness to an ideal solution and has been successfully applied in domains such as service quality assessment [30], technology selection [31], and educational tool evaluation [32].

Combining ISO/IEC 25010 with MCDM techniques like TOPSIS enables both technical and user-centric evaluation of educational tools. For example, Akargöl *et al*. [33] showed how TOPSIS could assist in the selection of e-learning platforms by employing weighted quality criteria. Despite its promise, few studies apply this combined approach in the context of AI-driven educational technologies, and even fewer incorporate teachers' and students' perspectives into the weighting of evaluation criteria. As Mishra and Koehler [34] emphasize in their Technological Pedagogical Content Knowledge (TPACK) framework, aligning technology with pedagogical goals and user needs is critical. This study addresses the research gap by integrating ISO/IEC 25010 and TOPSIS with stakeholder-weighted criteria, offering a holistic, replicable model for evaluating AIMAs in English-medium university instruction.

### *E. Summary and Research Significance*

This literature review reveals four key gaps in current research:

- A lack of inclusive technological support for linguistically diverse students in EMI settings
- Insufficient empirical evaluation of AI transcription tools in educational environments
- A lack of comprehensive comparative studies on AIMAs
- Minimal integration of standardized software quality frameworks and stakeholder-weighted MCDM models in the assessment of AI tools

To address these gaps, this study conducts a stakeholder-inclusive and comparative evaluation of ten AIMAs for EMI university lectures. It employs ISO/IEC 25010 and TOPSIS to ensure methodological rigor and a comprehensive quality assessment. The study aims to provide insights for higher education institutions aiming to enhance lecture comprehension, engagement, and accessibility for diverse student populations through the adoption of AI solutions.

## III. Methodology

This study used a systematic comparative evaluation design and a quantitative research methodology. Structured questionnaires, which produced quantitative data on participants' assessments of each AIMA across several evaluation criteria, were the main tool used to collect data. The within-subjects design enabled direct comparison of multiple tools by the same group of participants, providing a rigorous basis for comparative evaluation.

### *A. AIMA Tool Selection*

Ten AIMAs were chosen for evaluation based on their prominence in the market, applicability to educational contexts, and diversity of features as of May 2024. The selection process incorporated market research (e.g., comparative review blogs like [39]), a review of recent relevant literature, and consultation with educational technology experts from The Chinese University of Hong Kong (CUHK). Priority was given to tools with notable user adoption, AI functionalities (such as real-time transcription and automated note generation), and relevance to academic contexts. The final sample comprised Fireflies.ai, Grain, Krisp, Lark Minutes, Microsoft Teams, Nyota, Otter.ai, Reflect, Tl;dv, and Zoom. These tools represent a range of platforms developed by both major technology companies and startups. Detailed information on each tool can be found in Table A1.

### *B. Evaluation Framework*

The evaluation framework was grounded in the AdEQUATE Software Quality Evaluation Model [11], which is itself derived from the internationally recognized ISO/IEC 25010 standard for software quality. To better reflect educational contexts, the framework was expanded based on expert input from educational technologists to include two additional criteria: sustainability and scalability. In total, nine criteria were used to evaluate each AIMA: functional suitability, performance efficiency, compatibility, usability, reliability, security, satisfaction, sustainability, and scalability. Each criterion was operationalized through multi-item scales, adapted to reflect pedagogical requirements, and applied consistently across all tools for comparative purposes. Table A2 contains the definitions and measurement items for each criterion.

### *C. Application of the TOPSIS Method*

To synthesize both the performance ratings and the weighted stakeholder priorities into a final ranking, the TOPSIS method was applied. This approach produces a ranked list of AIMAs based on their relative closeness to the ideal solution [12]. The evaluation process began by constructing a decision matrix with the AIMAs as alternatives and the nine evaluation criteria as attributes. The matrix was normalized to allow for comparability across criteria. Importance ratings from participants were then applied as weights, yielding a weighted normalized matrix. The weight vector used was the arithmetic mean of the importance ratings that students and teachers submitted for each evaluation criterion. Next, the positive ideal solution (PIS) and negative ideal solution (NIS) were identified for each criterion. Subsequently, the Euclidean distance of each AIMA from both the PIS and NIS was computed. Finally, the relative closeness of each AIMA to the ideal solution, referred to as the TOPSIS score, was used to determine the performance ranking of the AIMAs.

### *D. Participants*

A total of twenty participants were recruited for the study, including twelve undergraduate students and eight teaching staff members. A snowball sampling strategy was initiated within the Faculty of Social Science and subsequently expanded to include a variety of faculties across the university. This approach was chosen because it enabled the efficient identification of both students and teaching staff from various faculties, including individuals who might be hard to reach or underrepresented through random sampling methods [40]. Giray [41] also used snowball sampling to assess student satisfaction with e-learning during the COVID-19 pandemic.

All participants were affiliated with CUHK, aged eighteen or older, free from cognitive impairments, and capable of understanding English. The student group included undergraduates from seven faculties: Engineering, Science, Medicine, Social Science, Business Administration, Law, and Arts, with most in their second year and with a balanced gender distribution. The teaching staff who participated were from the Faculties of Arts and Social Science, comprising both early-career and experienced educators, with an age range of 25 to 61. This diverse sample provided balanced perspectives from both learners and instructors.

### *E. Experimental Design and Procedures*

The evaluation took place between May and June 2024 using a within-subjects design, whereby each participant assessed all ten AIMAs. This approach enabled direct comparison of the tools and minimized variability arising from individual participant differences. To reduce potential order effects and participant bias, the presentation order of the AIMAs was counterbalanced across sessions, and participants received standardized instructions during the orientation phase.

The research procedure consisted of 4 stages. First, an orientation session was held to introduce participants to the study objectives, the nine evaluation criteria, and the AIMAs included in the assessment. Next, participants engaged in hands-on interaction with each AIMA in a simulated lecture environment, focusing on features such as real-time transcription, captioning, and notetaking. Immediately following each interaction, participants completed a structured questionnaire evaluating the AIMA according to the nine criteria.

After all tools had been assessed, participants completed a final assessment in which they provided ratings of the importance of each evaluation criterion for AIMA selection in academic contexts. Each session lasted approximately two hours, with breaks incorporated to minimize participant fatigue and maintain data quality. Participants tested the entry-level paid subscription for all AIMAs, which reflects the most economical premium options commonly adopted in educational settings. An exception was Microsoft Teams, which was evaluated using its free version and promotional materials due to licensing limitations.

### *F. Data Collection Instruments*

Two sets of questionnaires supported data collection. The AIMA Evaluation Questionnaire was given after each tool trial, measuring participants' views on the nine criteria with a 5-point Likert scale (1 = "Strongly Disagree," 5 = "Strongly Agree"). The items were primarily adapted from the AdEQUATE model and refined for relevance to educational technology use. At the end of each session, participants completed the Criteria Importance Questionnaire, which asked them to rate the importance of each criterion for selecting AIMAs in academic settings, also using a 5-point Likert scale (1 = "Not at All Important," 5 = "Very Important"). These importance ratings were subsequently used to determine the weights in the TOPSIS analysis. Both instruments underwent pilot testing to confirm their clarity, reliability, and contextual appropriateness.

### *G. Data Analysis*

Data analysis included several steps. Mean scores and standard deviations for each AIMA and criterion were calculated. Independent-samples t-tests were used to compare the importance ratings assigned by student and teacher groups. The TOPSIS method was then applied to synthesize performance data and importance weights, resulting in a final ranking of AIMAs that reflected both technical quality and stakeholder priorities. All statistical analyses used a significance threshold of $p < 0.05$.

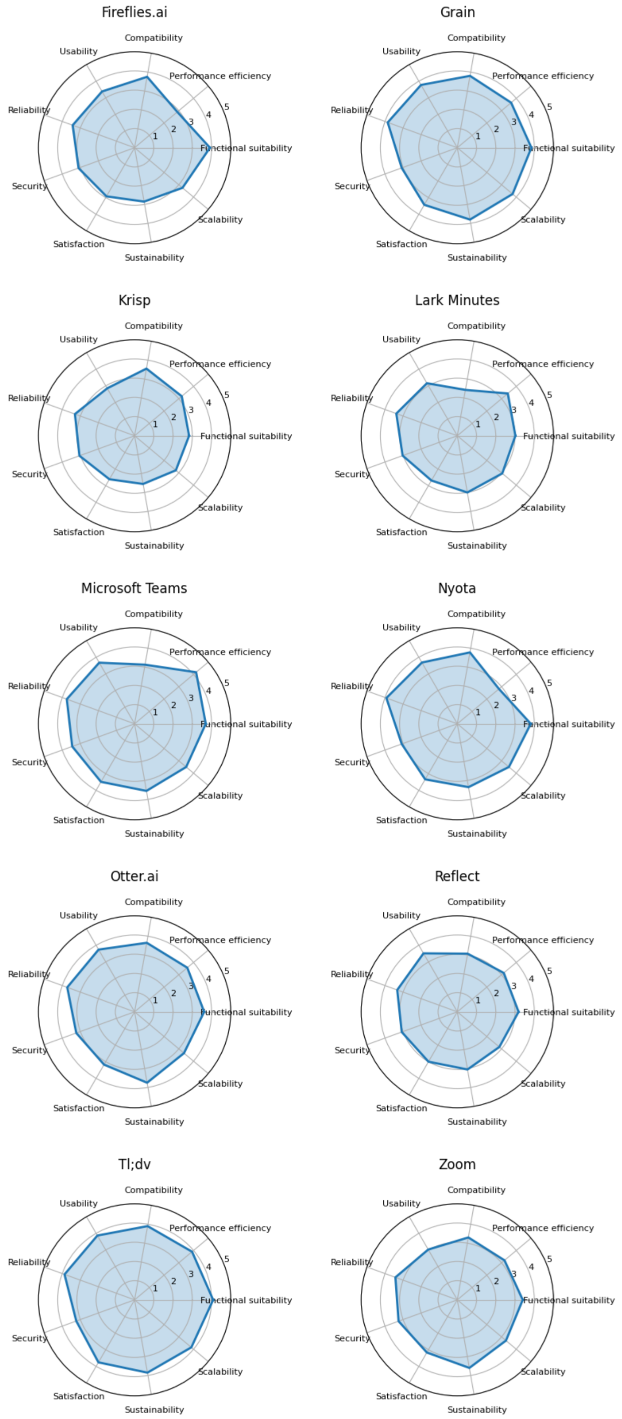


Fig. 1. Comparative radar charts of ten AIMAs across nine software quality criteria.

## IV. Results

### *A. Overall Evaluation of AIMAs*

As shown in Table 2, Tl;dv achieved the highest mean scores among all evaluated AIMAs across six critical software quality criteria: functional suitability ($M = 4.08$, $SD = 0.71$), compatibility ($M = 3.90$, $SD = 0.60$), usability ($M = 3.87$, $SD = 0.56$), satisfaction ($M = 3.76$, $SD = 0.91$), sustainability ($M = 3.85$, $SD = 0.81$), and scalability ($M =$

3.85, *SD* = 0.75). Microsoft Teams excelled in performance efficiency ($M = 4.17$, $SD = 0.59$) and security ($M = 3.45$, $SD = 0.48$), while Nyota ranked highest in reliability ($M = 3.92$, $SD = 0.63$). Standard deviations ranged from 0.45 to 1.20 across tools and criteria, reflecting moderate variability in participant assessments. This variation suggests a reasonable level of consensus among participants, while also indicating individual differences in user experience.

Table 2. Mean scores (standard deviations) for each AIMA across nine software quality criteria

| AIMA | Functional Suitability | Performance Efficiency | Compatibility | Usability | Reliability | Security | Satisfaction | Sustainability | Scalability |
|---|---|---|---|---|---|---|---|---|---|
| Fireflies.ai | 3.92 (0.55) | 2.90 (1.20) | 3.75 (0.68) | 3.38 (0.80) | 3.42 (1.00) | 3.10 (0.62) | 2.92 (1.11) | 2.85 (1.09) | 3.25 (0.91) |
| Grain | 3.85 (0.93) | 3.65 (0.83) | 3.80 (0.88) | 3.77 (0.66) | 3.85 (0.90) | 3.08 (0.67) | 3.43 (1.05) | 3.80 (0.70) | 3.75 (0.85) |
| Krisp | 2.83 (0.80) | 3.20 (0.95) | 3.55 (1.04) | 2.85 (0.95) | 3.30 (0.98) | 3.05 (0.81) | 2.62 (0.91) | 2.55 (1.00) | 2.80 (1.15) |
| Lark Minutes | 3.02 (1.12) | 3.42 (0.86) | 2.42 (0.86) | 3.16 (0.82) | 3.38 (1.09) | 3.02 (0.72) | 2.69 (1.09) | 3.00 (1.08) | 3.05 (1.05) |
| Microsoft Teams | 3.70 (0.63) | **4.17 (0.59)** | 3.12 (1.02) | 3.67 (0.45) | 3.75 (0.88) | **3.45 (0.48)** | 3.49 (0.68) | 3.55 (0.94) | 3.50 (0.76) |
| Nyota | 3.82 (0.83) | 2.83 (0.88) | 3.77 (0.94) | 3.68 (0.66) | **3.92 (0.63)** | 3.08 (0.77) | 3.34 (0.82) | 3.35 (0.88) | 3.50 (0.83) |
| Otter.ai | 3.62 (0.87) | 3.58 (1.10) | 3.65 (0.67) | 3.74 (0.65) | 3.73 (0.90) | 3.23 (0.77) | 3.17 (0.89) | 3.75 (1.02) | 3.35 (0.81) |
| Reflect | 3.18 (0.89) | 3.15 (1.04) | 3.08 (1.15) | 3.52 (0.60) | 3.33 (0.92) | 3.08 (0.69) | 3.00 (1.07) | 3.05 (1.19) | 2.85 (0.99) |
| Tl;dv | **4.08 (0.71)** | 3.90 (0.93) | **3.90 (0.60)** | **3.87 (0.56)** | 3.88 (0.94) | 3.23 (0.70) | **3.76 (0.91)** | **3.85 (0.81)** | **3.85 (0.75)** |
| Zoom | 3.40 (0.95) | 3.20 (1.14) | 3.30 (1.07) | 3.02 (0.79) | 3.42 (0.89) | 3.25 (0.73) | 3.17 (1.14) | 3.60 (0.99) | 3.30 (1.08) |

*Note:* Scores represent means, with standard deviations in parentheses. Bolded values indicate the highest mean score. All items rated on a 5-point scale (1 = Strongly Disagree, 5 = Strongly Agree).

To provide a visual summary of these results, Fig. 1 displays radar charts that illustrate the comparative performance profiles of each AIMA across the nine criteria. Tl;dv appears the most well-rounded, with high scores throughout. Microsoft Teams seems strongest in performance efficiency and security, while Nyota stands out in reliability. In contrast, some tools show more limited profiles, with scores clustered near the center. Overall, these charts suggest that some AIMAs offer broad strengths, while others excel in specific areas.

### *B. Stakeholder Importance Ratings for Evaluation Criteria*

Table 3 presents the mean importance ratings assigned by teachers and students. Functional suitability emerged as the highest priority overall ($M = 4.55$, $SD = 0.51$), followed by satisfaction ($M = 4.50$, $SD = 0.69$), performance efficiency ($M = 4.45$, $SD = 0.69$), and usability ($M = 4.45$, $SD = 0.69$).

Significant differences were observed between teachers and students for performance efficiency ($t = -2.93$, $p = 0.009$) and security ($t = -4.27$, $p < 0.001$), with teachers assigning greater importance to both criteria (see Table 4). No significant differences were found for the remaining criteria. When considering all criteria together, teachers rated them higher on average than students (teachers: $M = 4.49$; students: $M = 3.87$; $p = 0.0079$).

Table 3. Mean importance ratings (standard deviations) for software quality criteria by teachers and students

| Criterion | Teacher Rating | Student Rating | Total Rating |
|---|---|---|---|
| Functional Suitability | 4.75 (0.46) | 4.42 (0.51) | 4.55 (0.51) |
| Performance Efficiency | 4.88 (0.35) | 4.17 (0.72) | 4.45 (0.69) |
| Compatibility | 4.13 (1.36) | 3.83 (0.83) | 3.95 (1.05) |
| Usability | 4.75 (0.46) | 4.25 (0.75) | 4.45 (0.69) |
| Reliability | 4.50 (0.53) | 4.25 (0.62) | 4.35 (0.59) |
| Security | 4.63 (0.74) | 3.00 (0.95) | 3.65 (1.18) |
| Satisfaction | 4.63 (0.52) | 4.42 (0.79) | 4.50 (0.69) |
| Sustainability | 4.13 (1.13) | 3.25 (0.97) | 3.60 (1.10) |
| Scalability | 4.00 (0.76) | 3.25 (0.97) | 3.55 (0.94) |

*Note:* Scores represent means, with standard deviations in parentheses. Items rated on a 5-point scale (1 = Not Important, 5 = Extremely Important).

### *C. Weighted Ranking of AIMAs Using TOPSIS*

As displayed in Table 5, Tl;dv emerged as the top-performing AIMA (TOPSIS score: 0.90), followed by Grain (0.77) and Microsoft Teams Premium (0.72). Otter.ai (0.66) and Nyota (0.58) completed the top 5. This ranking indicates that Tl;dv consistently excelled in most evaluation criteria, aligning closely with teacher and student expectations. Grain and Microsoft Teams also performed strongly as alternatives.

Table 4. Results of Independent-samples t-tests comparing teacher and student ratings across evaluation criteria

| Criterion | t-statistic | p-value | Significance |
|---|---|---|---|
| Functional Suitability | −1.508 | 0.151 | Not Significant |
| Performance Efficiency | −2.927 | 0.009 | Significant |
| Compatibility | −0.543 | 0.598 | Not Significant |
| Usability | −1.836 | 0.083 | Not Significant |
| Reliability | −0.959 | 0.351 | Not Significant |
| Security | −4.268 | 0.000 | Significant |
| Satisfaction | −0.711 | 0.486 | Not Significant |
| Sustainability | −1.801 | 0.094 | Not Significant |
| Scalability | −1.942 | 0.068 | Not Significant |

*Note:* Significance defined as $p < 0.05$.

Table 5. Weighted TOPSIS scores and ranks for the top 5 AIMAs

| AIMA | TOPSIS Score | Rank |
|---|---|---|
| Tl;dv | 0.90 | 1 |
| Grain | 0.77 | 2 |
| Microsoft Teams | 0.72 | 3 |
| Otter.ai | 0.66 | 4 |
| Nyota | 0.58 | 5 |

*Note:* Scores reflect the relative closeness to the ideal solutions as calculated using the TOPSIS method. Only the top 5 tools are shown.

## V. DISCUSSION

The results of this study indicate that Tl;dv, Grain, and Microsoft Teams ranked highest overall in the multi-criteria evaluation that involve technical quality and stakeholder priorities. This finding partially aligns with [6], who found that Otter.ai improved students' academic confidence but did not always enhance classroom dynamics or instructor clarity. In contrast, our results suggest that Tl;dv and Grain perform better in multiple aspects, e.g., usability, compatibility, and satisfaction. When compared to Khoo *et al*. [27], who noted that Whisper and TurboScribe outperformed Otter.ai in accuracy and that Whisper's local transcription preserves user privacy, our study found that Microsoft Teams was rated highest for security and performance efficiency. These diverse stakeholder perspectives also support the recommendations of Asthana *et al*. [35] and Cabrero-Daniel *et al*. [20], who advocate for the customization of AIMAs to address varying user needs.

### A. Theoretical Interpretation of Findings

This study conducted a systematic evaluation of ten AIMAs using a multi-criteria framework based on ISO/IEC 25010 and TOPSIS, thereby showing how each tool scored on every quality criterion (RQ1) and which tool emerged as the overall leader when stakeholder weights were applied (RQ2). As a result, Tl;dv consistently outperformed other tools across all nine software quality criteria, reflecting its strong alignment with both student and teacher preferences as the top performer. Regarding RQ3, which asks how teacher and student priorities diverge, teachers assigned greater importance to performance efficiency and security than students, illustrating distinct stakeholder perspectives in the adoption of educational technology. Their overall higher ratings also point to a more cautious, system-focused approach to AIMAs for classroom use.

In relation to RQ2, established theoretical models help explain why certain AIMAs performed better than others when user priorities were considered. The Technology Acceptance Model (TAM) is particularly relevant, as it identifies perceived usefulness and ease of use as central drivers of technology adoption [22]. Tl;dv's high scores in functional suitability, usability, and satisfaction closely aligns with these TAM constructs, which helps explain its top ranking in the stakeholder-weighted analysis. The Technological Pedagogical Content Knowledge (TPACK) framework also offers a relevant lens, emphasizing the integration of technological, pedagogical, and content knowledge [34]. The varied performance of AIMAs on different evaluation criteria highlights differing degrees of fit with TPACK principles. Tools such as Tl;dv and Grain demonstrate a better balance of pedagogical utility and technological strength, while lower-ranked options appear to focus more narrowly on technical capabilities at the expense of pedagogical relevance.

To address RQ3, the study also examined how teachers and students differ in their evaluation priorities. Teachers in this study prioritized security and performance efficiency. This result echoes the findings of previous research on barriers to digital transformation in higher education institutions, which highlighted concerns about data protection and institutional risk [42]. This pattern also fits with Diffusion of Innovation theory, which notes that compatibility and perceived long-term benefit (relative advantage) often matter more to those responsible for implementing new systems [43]. Following the same logic, it is reasonable for students to prioritize immediate functionality and ease of use to support their coursework. While such contrasts are not surprising, they do highlight the need for evaluation frameworks that reflect more than one point of view.

Finally, the high importance placed on functional suitability and satisfaction across both groups supports a key conclusion related to RQ1. Effective AI implementation in education requires both technical reliability and a positive user experience, as also emphasized in [44]. The performance differences observed among the ten AIMAs point to ongoing challenges in adapting commercial tools for academic use. As noted by Kuhn *et al*. [19], some AIMAs lack sufficient alignment with pedagogical goals. These findings highlight the value of using a stakeholder-weighted and standards-based muti-criteria decision-making model to support informed, inclusive, and context-specific adoption of AI in higher education.

In summary, the genuine contribution of this study lies in its application of a holistic, multi-criteria decision-making framework (ISO/IEC 25010 and TOPSIS) which included the viewpoints of both teachers and students. Unlike prior research that typically addresses a single stakeholder group or limited software dimensions, our approach provides a comprehensive and nuanced evaluation that can guide institutional adoption of AIMAs. This study closes a significant gap in the literature and provides a workable, replicable model for evidence-based technology selection in higher education in the context of English-medium university instruction.

### B. Practical Implications

These findings highlight several practical implications for stakeholders in educational technology. For educators and learners, high-performing tools like Tl;dv and Grain show great promise for teacher and student support of classroom activities, especially in EMI contexts, with features such as live transcription, lecture recording, and content summarization. Their overall performance across core quality dimensions suggests their suitability for general academic use. At the same time, performance variance, e.g., Microsoft Teams scoring higher in efficiency and security, suggests that certain tools might be more appropriate for particular educational requirements or institutional settings. These results can help faculty and students make informed choices when selecting AIMA tools that best fit their requirements. Teachers and students are also encouraged to share feedback with their IT departments on issues such as usability, data privacy, and reliability regularly.

For developers and vendors, the results identify targeted areas for improvement. Developers should prioritize quality dimensions most valued by educational stakeholders to increase adoption. For example, enhancing aspects like functional suitability could help improve the tool's educational viability.

For higher education institutions, the study underscores the importance of adopting systematic, evidence-based evaluation processes grounded in transparent and stakeholder-relevant criteria. Rather than relying on ad hoc testing or marketing-driven decisions, institutions may consider updating their procurement policies to incorporate both technical standards and the distinct priorities of user groups such as faculty and students. As an additional contribution, this study provides a replicable evaluation framework that integrates ISO/IEC 25010 software quality standards with the TOPSIS-based MCDM method. Institutions are encouraged to periodically reapply this framework within their own context. This act can ensure that AIMA evaluations remain up to date and aligned with evolving technological capabilities and stakeholder expectations.

### C. Policy and Sustainability Considerations

The present findings suggest three considerations for policy and the sustainable adoption of AIMA technologies in higher education. First, it is recommended that educational institutions adopt evidence-based procurement procedures that take into account the opinions of both students and instructors. Successful integration of educational technology in classes requires the acceptance and positive attitudes of both teachers and students [13]. Yet, the needs of teachers and students often may not be adequately reflected in traditional procurement methods such as marketing research and ad hoc trials alone. Therefore, before purchasing, institutions may request that vendors demonstrate how their software meets ISO/IEC 25010 standards.

Second, in order to handle privacy concerns brought about by AI-powered transcription tools, data-governance frameworks need to be updated. AI technologies have the potential to greatly improve teaching and learning, but they may also raise issues with data security and student participation in the classroom [45]. For instance, students may participate less in class if they believe that their input is being recorded and may be shared without their permission. Users may want reassurance that their data is handled legally and strictly for its intended purpose, as [46] emphasizes. Therefore, to reduce privacy risks and promote trust, precise rules regarding data retention, informed consent procedures, and thorough audit trails should be put in place.

Lastly, continuous assessment is essential for sustainable technology adoption. As user expectations and technological capabilities change, universities should establish systematic procedures for routinely reevaluating AIMA tools. Maintaining alignment with institutional and pedagogical needs will be made easier with the regular use of the proposed stakeholder-weighted ISO/IEC 25010 + TOPSIS evaluation framework in this paper.

### D. Limitations and Future Research

A notable strength of this study lies in its attempt to bring together technical evaluation with the perspectives of both students and teachers. By applying internationally recognized ISO/IEC 25010 frameworks alongside the TOPSIS method, the analysis goes beyond basic feature comparisons and instead focuses on evaluating how well these tools meet the needs of real users in educational settings.

However, several limitations warrant consideration. The first limitation concerns sample size. While participants represented diverse disciplines and backgrounds, the modest number limits the generalizability of the findings and suggests that rankings should be interpreted with caution until validated across broader populations. Future research would benefit from larger, multi-institutional samples that examine the effects of AIMAs on student learning and participation in varied settings. Another limitation of this study is that we did not collect information on participants' prior experience with AIMAs. Although all participants received a standardized introduction to each tool and sufficient time to trial them before evaluation, it is still possible that varying levels of prior exposure or familiarity with AIMAs may have influenced their preferences and assessments. Future research should investigate the impact of users' previous experience with AIMAs on their evaluation and decision-making processes. A third limitation involves the focus of this evaluation. The study centered on technical and usability criteria, while providing less attention to how these tools support pedagogical integration or influence learning outcomes. Further assessment of AIMAs' impact on student comprehension, engagement, and achievement is needed. Longitudinal studies tracking both instructional practices and student experiences over time would also offer valuable insights into sustained adoption and educational benefit. The timing of the evaluation also presents challenges for interpreting the results. Data collection of this study took place during a brief window in mid-2024, yet AI technologies evolve rapidly. The evaluation or ranking presented here should be considered provisional. The dynamic nature of AI tools means that system capabilities and user priorities may shift significantly over short periods. As such, future research could make good use of the demonstrated MCDM framework to develop dynamic evaluation tools, e.g., automated dashboards or online platforms, that allow for ongoing, user-driven assessment and facilitate timely updates to institutional technology decision-making. Finally, because teacher and student importance ratings were aggregated, the analysis may mask subgroup differences. Future studies should test separate weighting scenarios to determine how distinct stakeholder priorities influence AIMA rankings.

## VI. Conclusion

This research contributes a systematic and replicable framework for evaluating AIMAs in higher education, explicitly combining technical standards with stakeholder-driven priorities. With respect to RQ1, the study discovered notable variations in the performance of ten AIMAs when assessed against ISO/IEC 25010 criteria. Tl;dv received the highest mean scores across key criteria like functional suitability, compatibility, usability, satisfaction, and scalability. In contrast, Microsoft Teams performed the best in performance efficiency and security. In response to RQ2, the stakeholder-weighted TOPSIS analysis revealed that Tl;dv ranked highest overall (TOPSIS score = 0.90), followed by Grain (0.77) and Microsoft Teams (0.72). For RQ3, the findings uncovered significant differences between teachers and students in their evaluation priorities. Teachers tended to give more weight to performance efficiency and security than students.

In practical terms, these findings offer evidence-based recommendations for teachers and students incorporating AIMAs into EMI classrooms, institutions looking to implement or update digital meeting tools, and developers hoping to address educational priorities in subsequent iterations. However, this study is subject to several limitations, including a relatively small and institutionally localized sample size and the rapidly evolving nature of AI technology, which may affect the long-term generalizability of these results. Although the current study has identified the AIMAs' current leaders and outlined their distinct strengths, it also emphasizes the necessity of ongoing, stakeholder-informed evaluation as user needs and technology change. The integration of MCDM frameworks like ISO/IEC 25010 and TOPSIS, alongside attention to diverse user perspectives, offers a robust model for ongoing, sustainable technology adoption in higher education.

## APPENDIX

Table A1. Overview of selected AIMA tools

| Tool | User Metrics (Date & Source Type) | Focus | HQ / Origin | Native LMS Integration | Source |
|---|---|---|---|---|---|
| Fireflies.ai | 20M users (Apr 2025, business news) | Business – AI meeting notes, CRM automations | USA | No | [47, 48] |
| Grain | NPD – marketing claim "31K teams" (Jun 2025, official site) | Business – AI meeting insights for sales & research | USA | No | [49] |
| Krisp | 200M devices (Jun 2025, official blog) | Business / Call-centers – Voice/Accent AI, agent-assist | USA | No | [50, 51] |
| Lark Minutes (Feishu) | 4.5M DAU (Nov 2021, business news) | Business – Collaboration suite | Singapore / China | No | [52, 53] |
| Microsoft Teams | 320M MAU (Oct 2023, earnings conference) | Business + Education – Collaboration platform | USA | Yes (LTI Meetings/Classes for Moodle and Backboard, etc.) | [54, 55] |
| Nyota | NPD – marketing claim "Trusted by 1000s of teams" (Jun 2025, official site) | Business – AI notes for sales & projects | UK | No | [56] |
| Otter.ai | 25M users (Mar 2025, business news) | Business + Education – AI transcription & meeting agents | USA | No | [57, 58] |
| Reflect | NPD – no statistics publicly shared | Personal / Knowledge-work – Personal note-taking with AI | USA | No | [59] |
| Tl;dv | 900K users (Feb 2023, third-party review) | Business – Async meeting recorder & summaries | Germany | No | [60, 61] |
| Zoom | 300M DAU (Jun 2025, third-party blog) | Business + Education – Video conferencing platform | USA | Yes (LTI Pro for Moodle, Blackboard, etc.) | [62, 63] |

*Notes:* All user numbers are vendor-reported and unaudited; Devices count installations, not unique people; CRM = Customer Relationship Management; DAU = Daily Active Users; HQ = Headquarter; LMS = Learning Management System; LTI = Learning Tools Interoperability; MAU = Monthly Active Users; NPD = Not Publicly Disclosed.

Table A2. List of constructs and corresponding items

| Construct | Definition | Item |
|---|---|---|
| Functional suitability | Extent to which the tool provides necessary features to support lecture engagement and comprehension. | ○ The system has functionalities that I would expect it to have.<br>○ The system provides the correct results with the necessary degree of precision.<br>○ The system functionalities facilitate the fulfillment of my tasks. |
| Performance efficiency | Speed and responsiveness in real-time scenarios. | ○ The system responds quickly.<br>○ I can complete my tasks quickly. |
| Compatibility | Ability to interoperate with other platforms and share system resources. | ○ The system executes its functionalities efficiently while sharing its environment and resources with other products (e.g., Zoom and Panopto).<br>○ The system allows the exchange of information with other systems when necessary. |
| Usability | Ease of use, user interface integration, and accessibility. | ○ I think that the various system functions are well integrated.<br>○ Learning to use the system is easy for me.<br>○ Tasks can be performed in a straightforward manner with this system.<br>○ The system protects me from making errors.<br>○ The system interface is pleasant.<br>○ I can use the system despite my special needs. |
| Reliability | Consistency and stability of performance. | ○ The systems always behaves as expected.<br>○ The system never stops unexpectedly. |
| Security | Protection of user data and prevention of unauthorized access. | ○ I am certain that the system data are available only to authorized people.<br>○ I am certain that the system blocks all unauthorized access to the program or its data. |
| Satisfaction | Overall user experience and likelihood of future use. | ○ In general, I think the system is useful in my job.<br>○ I would recommend this system to my colleagues / classmates<br>○ I think I would use this system frequently.<br>○ I felt very confident using the system.<br>○ Interacting with the system is usually compensating.<br>○ I feel comfortable using the system. |
| Sustainability | Long-term viability without additional resource investment. | ○ I believe that this software would continue to bring impact to students even if no extra resources are given. |
| Scalability | Suitability for broader or alternative use cases beyond initial scope. | ○ I believe that this software would be suitable for use cases beyond what it was originally designed for. |

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Alex Pak Ki Kwok was responsible for the conceptualization, methodology development, formal analysis, validation, original draft writing, review and editing. Yao Hing Wong contributed to the methodology development, investigation, and review and editing of the manuscript. Kwong-Cheong Wong and Chee Hon Chan provided research guidance and contributed to the review and editing of the manuscript. All authors have approved the final version of the manuscript.

## FUNDING

This research was funded by the Funding Scheme to Enhance Student Engagement and Address Student Learning Needs (Supported by Teaching Development and Language Enhancement Grant 2022-25), grant number 4171072.

## ACKNOWLEDGMENT

The authors wish to thank all the participants for involving in the experiment.

## REFERENCES

[1] J. Dearden, *English as a Medium of Instruction—A Growing Global Phenomenon*, British Council, 2014.
[2] F. Helm, "EMI, internationalisation, and the digital," *International Journal of Bilingual Education and Bilingualism,* vol. 23, no. 3, pp. 314–325, 2020.
[3] L. Yuan, F. Fang, and G. Hu, "Are students prepared and supported for English medium instruction in Chinese higher education to promote educational equality?" *System,* vol. 125, 103423, 2024.
[4] Q. Zheng and T.-H. Choi, "English-medium instruction as an internationalisation strategy at a second-tier Chinese University: instructors' challenges and their shaping factors," *Asian-Pacific Journal of Second and Foreign Language Education,* vol. 9, no. 1, p. 76, 2024.
[5] K. J. Williams, M. L. R., F. Anna-Mária, M. Jeremy, and S. and Vaughn, "Student engagement among high school English learners with reading comprehension difficulties," *School Psychology Review,* vol. 52, no. 1, pp. 38–56, 2023.
[6] A. P. K. Kwok, Y. H. Wong, K. C. Wong, and C. H. Chan, "AI meeting assistants in English-medium university lectures in Hong Kong, China: A double-edged sword for student perception," *International Journal of Information and Education Technology,* vol. 14, no. 9, 2024.
[7] S. Malakul and I. Park, "The effects of using an auto-subtitle system in educational videos to facilitate learning for secondary school students: Learning comprehension, cognitive load, and satisfaction," *Smart Learning Environments,* vol. 10, no. 1, p. 4, 2023.
[8] R. Kelly, "Survey: 86% of students already use ai in their studies," *Campus Technology*, 2024.
[9] TechPipeline, "Survey of U.S. higher education faculty 2024: Contributions of content & design to AI applications," *TechPipeline*, 2024.
[10] Data Bridge Market Research, "Global AI meeting assistants market, by type, application—industry trends and forecast to 2031," *Data Bridge Market Research*, Fremont, CA, USA, 350, 2024.
[11] J. Alves, C. G. Wangenheim, T. Lacerda, A. Savaris, and A. Wangenheim, *AdEQUATE Software Quality Evaluation Model v1.0*. 2015.
[12] C.-L. Hwang and K. Yoon, "Methods for multiple attribute decision making," *Multiple Attribute Decision Making: Methods and Applications A State-of-the-Art Survey*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 58–191.
[13] Q. Li, "Student and teacher views about technology: A tale of two cities?" *Journal of research on Technology in Education,* vol. 39, no. 4, pp. 377–397, 2007.
[14] M. Li and L. Pei, "Exploring challenges in academic language-related skills of EFL learners in Chinese EMI settings," *Acta Psychol (Amst),* vol. 247, 104309, 2024.
[15] L. Xuan Mai, L. K. Ngoc, and L. T. Thao, "Factors hindering student participation in English-speaking classes: Student and lecturer perceptions," *SAGE Open,* vol. 14, no. 3, 2024.
[16] J. J. Chen and C. and Perez, "Enhancing assessment and personalized learning through artificial intelligence," *Childhood Education,* vol. 99, no. 6, pp. 72–79, 2023.
[17] S. F. Ahmad, M. M. Alam, M. K. Rahmat, M. S. Mubarik, and S. I. Hyder, "Academic and administrative role of artificial intelligence in education," *Sustainability,* vol. 14, no. 3, p. 1101, 2022.
[18] Y. Song, L. R. Weisberg, S. Zhang, X. Tian, K. E. Boyer, and M. Israel, "A framework for inclusive AI learning design for diverse learners," *Computers and Education: Artificial Intelligence,* vol. 6, 100212, 2024.
[19] K. Kuhn, V. Kersken, B. Reuter, N. Egger, and G. Zimmermann, "Measuring the accuracy of automatic speech recognition solutions," *ACM Trans. Access. Comput.,* vol. 16, no. 4, 2024.
[20] B. Cabrero-Daniel, T. Herda, V. Pichler, and M. Eder, "Exploring human-AI collaboration in agile: Customised LLM meeting assistants," in *Proc. International Conference on Agile Software Development*, Springer Nature Switzerland Cham, 2024, pp. 163–178.
[21] H. S. Khoo, R. Li, G. A. Yee, and W. T. Li-Lian, "Transcription platforms: A comparative analysis of AI-assisted tools," presented at the Value-Based Healthcare Conference 2024, Singapore, 2025.
[22] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly,* vol. 13, no. 3, pp. 319–340, 1989.
[23] A. S. Al-Adwan, N. Li, A. Al-Adwan, G. A. Abbasi, N. A. Albelbisi, and A. Habibi, "'Extending the Technology Acceptance Model (TAM) to predict university students' intentions to use metaverse-based learning platforms'," *Education and Information Technologies,* vol. 28, no. 11, pp. 15381–15413, 2023.
[24] S. S. Jasrotia, A. P. K. Kwok, and S. Koul, "Exploring the impact of social relationships and system experiences on consumers' acceptance of social commerce in the fashion sector: an empirical study," *International Journal of Information Management Data Insights,* vol. 5, no. 1, 100333, 2025.
[25] L. Xue, A. M. Rashid, and S. Ouyang, "The Unified Theory of Acceptance and Use of Technology (UTAUT) in higher education: A systematic review," *SAGE Open,* vol. 14, no. 1, 21582440241229570, 2024.
[26] M. Falco and G. Robiolo, "Building a catalogue of ISO/IEC 25010 quality measures applied in an industrial context," *Journal of Physics: Conference Series,* vol. 1828, no. 1, 012077, 2021.
[27] J. P. Marroquin and C. R. Rodriguez, "Quality assessment model based on ISO/IEC 25010 for e-learning platforms in the cloud and based on laravel," in *Proc. 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2023, pp. 751–757.
[28] I. E. I. Fadhel, S. Z. B. S. Idrus, M. S. Y. Abdullah, A. A. E. A. Ibrahim, M. Omar, and A. Khred, "Towards development a novel framework of web-based systems quality engineering by the integration between information systems and software engineering theories: Context of higher education," *Journal of Physics: Conference Series*, IOP Publishing, vol. 1529, no. 2, 022005, 2020.
[29] S. Chakraborty, "TOPSIS and Modified TOPSIS: A comparative analysis," *Decision Analytics Journal,* vol. 2, 100021, 2022.
[30] C.-L. Lin, J.-J. Chen, and Y.-Y. Ma, "Ranking of service quality solution for blended design teaching using fuzzy ANP and TOPSIS in the post-COVID-19 era," *Mathematics,* vol. 11, no. 5, p. 1255, 2023.
[31] G. Boonsothonsatit, S. Vongbunyong, N. Chonsawat, and W. Chanpuypetch, "Development of a hybrid AHP-TOPSIS decision-making framework for technology selection in hospital medication dispensing processes," *IEEE Access,* vol. 12, pp. 2500–2516, 2024.
[32] S. Maitra, L. Sahoo, J. Lahiri Dey, and K. Tiwary, "Multi-criteria decision making and its application to online learning platform selection during the COVID-19 pandemic based on TOPSIS method," in *Real Life Applications of Multiple Criteria Decision Making Techniques in Fuzzy Domain*, L. Sahoo, T. Senapati, and R. R. Yager Eds., Singapore: Springer Nature Singapore, 2023, pp. 507–519.
[33] İ. Akargöl, İ. Karadağ, and Ö. F. Gürcan, "Selecting the optimal e-learning platform for universities: A pythagorean fuzzy AHP/TOPSIS evaluation," *The European Journal of Research and Development,* vol. 4, no. 2, pp. 19–34, 2024.
[34] P. Mishra and M. J. Koehler, "Technological pedagogical content knowledge: A framework for teacher knowledge," *Teachers College Record,* vol. 108, no. 6, pp. 1017–1054, 2006.
[35] S. Asthana, S. Hilleli, P. He, and A. Halfaker, "Summaries, highlights, and action items: Design, implementation and evaluation of an LLM-powered meeting recap system," in *Proc. the ACM on Human-Computer Interaction,* vol. 9, no. 2, pp. 1–29, 2025.
[36] L. Haliburton, N. Bartłomiejczyk, A. Schmidt, P. W. Woźniak, and J. Niess, "The walking talking stick: Understanding automated note-taking in walking meetings," in *Proc. the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–16.
[37] A. Herdiyanti, "The use of automatic ai-based notes and transcription services in qualitative research: Ethical and methodological concerns," in *Proc. the ALISE Annual Conference*, 2024,
[38] S. Son, J. Choi, S. Lee, J. Y. Song, and I. Shin, "It is okay to be distracted: how real-time transcriptions facilitate online meeting with distraction," in *Proc. the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–s19.
[39] R. Miguel, *The 8 Best AI Meeting Assistants in 2024*, Zapier, 2024.
[40] M. Törnävä, M. Palonen, E. Harju, T. Haapa, M.-L. Rissanen, and J. Kylmä, "Hard-to-reach and hidden groups in health-related research—A scoping review," *Journal of Advanced Nursing,* 2025.
[41] G. Giray, "An assessment of student satisfaction with e-learning: An empirical study with computer and software engineering undergraduate students in Turkey under pandemic conditions,"

*Education and Information Technologies,* vol. 26, no. 6, pp. 6651–6673, 2021.
[42] T. Gkrimpizi, V. Peristeras, and I. Magnisalis, "Classification of barriers to digital transformation in higher education institutions: Systematic literature review," *Education Sciences,* vol. 13, no. 7, p. 746, 2023.
[43] E. M. Rogers, A. Singhal, and M. M. Quinlan, "Diffusion of innovations," *An Integrated Approach to Communication Theory and Research*, Routledge, 2014, pp. 432–448.
[44] S. Ntoa, "Usability and user experience evaluation in intelligent environments: A review and reappraisal," *International Journal of Human–Computer Interaction,* vol. 41, no. 5, pp. 2829–2858, 2025.
[45] F. Filgueiras, "Artificial intelligence and education governance," *Education, Citizenship and Social Justice,* vol. 19, no. 3, pp. 349–361, 2024.
[46] M. I. Khalid, M. Ahmed, and J. Kim, "Enhancing data protection in dynamic consent management systems: Formalizing privacy and security definitions with differential privacy, decentralization, and zero-knowledge proofs," *Sensors,* vol. 23, no. 17, p. 7604, 2023.
[47] Fireflies AI, *Beyond Notetaking: Fireflies Offers 200+ Agentic AI Apps to Amplify Team Performance Across Every Department*, SAN FRANCISCO: Business Wire, 2025.
[48] Fireflies.ai., "Fireflies.ai: AI notetaker to transcribe, summarize & analyze meetings." Fireflies.ai Corp, 2025.
[49] Grain Intelligence Inc., *Grain: The AI Notetaker Built for Growing Teams*, 2025.
[50] Krisp Team, *Krisp Launches All-In-One Voice Productivity Platform for Contact Centers*, Berkeley, CA: Krisp Technologies, Inc., 2025.
[51] Krisp Technologies Inc., *Krisp: AI Meeting Assistant with Noise Cancellation, Transcription & More*, 2025.
[52] L. Emma, "Lark, ByteDance's Slack-like app, eyes $1 billion global revenue in five years," Shanghai: TechNode, 2021.
[53] Lark Technologies Inc., "Lark Minutes: AI meeting transcription, notes & summaries," *Lark Suite*, 2025.
[54] I. S. Brett *et al*., *Microsoft Fiscal Year 2024 First Quarter Earnings Conference Call*, Microsoft Corporation, 2023.
[55] Microsoft Support, "Use Microsoft Teams meetings in your LMS," *Microsoft Corporation,* 2025.
[56] Nyota A. I. Ltd., *Nyota.ai: The One AI Notetaker to Align Sales, Support, and Project Teams*, 2025
[57] M. Woodrow, *Otter.ai Breaks $100M ARR Barrier and Transforms Business Meetings, Launching Industry-First AI Meeting Agent Suite*, Mountain View, CA: Business Wire, 2025.
[58] Otter.ai Inc., *Otter.ai: The #1 AI Meeting Assistant for Transcription, Summaries & Action Items*, 2025.
[59] Reflect App LLC, *Reflect.app: A Privacy-Focused AI Notetaking & Knowledge Tool*, 2025.
[60] J. Woodwalker, "TL;DV: AI meeting recorder for summaries & transcripts," *AllThingsAI (Overview),* 2025.
[61] Tldv, *tl;dv: AI Notetaker for Zoom, Microsoft Teams & Google Meet*, 2025.
[62] K. Naveen, *Zoom User Statistics 2025—Market Share & Active Users*, Sacramento, CA: DemandSage, 2025.
[63] Zoom Support, "Configuring LTI pro," *Zoom Video Communications*, 2025.