Parrot: An Agentic Classroom AI

Kalena Dai [1] **Arya Sarukkai** [2]

]

Abstract

We introduce **Parrot**, an interpretable, multimodal AI agent designed to enhance real-time teaching and learning in classrooms. Parrot operates autonomously as both a curious student and an assistant lecturer, performing actions such as summarizing lecture content, detecting engagement via multimodal sentiment analysis, and generating context-aware questions. The system integrates Retrieval-Augmented Generation (RAG) grounded in curriculum materials, DeepPrivacy2 for real-time face anonymization, and adaptive learning capabilities. Each classroom instance locally adapts its strategies while contributing anonymized metadata to improve shared retrieval and prompt policies via federated collaboration. A dedicated Learner module continuously refines Parrot's retrieval logic and prompting behaviors, enabling long-term improvement without compromising privacy. We present results from simulated deployments and discuss how Parrot exemplifies agentic intelligence in education through adaptability, transparency, and trustworthy autonomy.

1. Introduction

Classrooms are dynamic, high-context environments that demand real-time awareness and pedagogical flexibility. Yet most educational AI tools remain static and opaque, functioning more as dashboards than active collaborators. In this work, we present Parrot, a multimodal classroom agent built on principles of agentic AI—autonomous systems that perceive, reason, and act in human-centered settings.

Parrot listens to lectures, monitors engagement using audiovisual and textual cues, summarizes instruction, and generates timely, pedagogically relevant questions. It emphasizes actionable interpretability, allowing teachers to understand and validate its outputs, and integrates DeepPrivacy2 to preserve student anonymity while maintaining sentiment fidelity.

The system adopts a dual-persona model: a Curious Student that asks clarifying or exploratory questions, and an Assistant Lecturer that answers in-context queries. These roles are supported by a RAG pipeline grounded in course materials, enabling both relevance and transparency. All components are coordinated through a central Controller and refined over time by a Learner module, which supports both local adaptation and federated knowledge-sharing across classrooms.

Through this architecture, Parrot embodies agentic traits—cooperation, contextual reasoning, and adaptability—making it a promising step toward AI systems that actively enhance human learning experiences.

Key contributions of this work include:

- A real-time, interpretable classroom AI agent that integrates sentiment analysis, curriculum-grounded RAG, and dual instructional personas.
- A privacy-preserving architecture powered by DeepPrivacy2, ensuring compliance with regulations while preserving emotional fidelity.
- A Learner framework for adaptive and federated refinement, enabling the system to evolve based on classroom feedback without compromising privacy.
- Empirical validation in simulated classroom settings, demonstrating engagement gains and alignment with teaching goals.

2. Related Work

Explainable AI in Education: Recent work emphasizes transparency in educational AI, advocating for *teacher-in-the-loop* systems (Khosravi et al., 2022; Holstein et al., 2018). Traditional intelligent tutors like AutoTutor (Graesser et al., 2005) laid important groundwork but often lacked interpretability. The U.S. Department of Education (U.S. Department of Education, Office of Educational

[1]Mission San Jose High School, Fremont CA, US [2]Saratoga High School, Saratoga CA, US. Correspondence to: Kalena Dai <kalenakkdai@gmail.com>.

Accepted at the ICML 2025 Workshop on Collaborative and Federated Agentic Workflows (CFAgentic@ICML'25), Vancouver, Canada. July 19, 2025. *Copyright 2025 by the author(s).*

Technology, 2023) highlights the need for educators to understand and override AI decisions. Parrot aligns with these principles by offering interpretable outputs—e.g., sentiment alerts include the model's reasoning based on facial or tonal cues, and answers are grounded in specific retrieved content.

Multimodal Sentiment & Engagement Analysis: Combining multiple data modalities improves affect detection in learning environments. Pan et al. (Pan et al., 2024) fused face, gait, and attention models to recognize classroom emotions with over 85% accuracy. Others (Whitehill et al., 2014; D'Mello & Kory, 2015) have shown that integrating facial, gestural, and vocal cues outperforms unimodal approaches. Parrot builds on these findings by using synchronized audio (tone), video (expressions), and textual (transcripts) signals to assess student engagement in real time.

Privacy-Preserving Facial Anonymization: With growing concerns over classroom surveillance, privacy-preserving techniques have become critical. DeepPrivacy2 (Hukkelås et al., 2023) is a GAN-based model that replaces faces with photorealistic surrogates while retaining emotional features like gaze and expression. Its compliance with FERPA (U.S. Department of Education, 2023) and similar policies makes it ideal for classroom deployment. Parrot integrates DeepPrivacy2 to ensure high-fidelity sentiment analysis while protecting identity.

AI Teaching Assistants and LLMs: LLMs such as GPT-4 have been applied to education as tutors and classroom assistants. Khan Academy's Khanmigo (Khan Academy, 2023) uses GPT-4 to scaffold problem-solving. Long et al. (Long et al., 2024) showed that GPT-4o could reliably label classroom transcripts for instructional feedback. Ensuring output alignment with course content remains a key challenge, which Retrieval-Augmented Generation (RAG) (Lewis et al., 2020) helps address by grounding answers in syllabus-aligned materials. Parrot adopts RAG to ensure responses are both accurate and explainable.

Multimodal Classroom Agents: Vision-based agents like VidAAS (Lee et al., 2024) use multimodal inputs to provide rubric-aligned instructional feedback. While Parrot focuses more on student engagement and instructional summarization, both systems emphasize *actionable interpretability*—ensuring AI outputs are usable and understandable by teachers.

Adaptive Learning and Federated Collaboration: Adaptive learning systems dynamically adjust content based on student behavior and have been shown to increase engagement and understanding (Brusilovsky & Millán, 2007). Parrot's *Learner* module similarly tunes its sentiment thresholds and question strategies based on feedback within each classroom session. On a broader scale, Parrot participates in *federated learning* (McMahan et al., 2017), aggregating anonymized insights (e.g., effective prompt patterns, high-yield RAG retrievals) from many classrooms without sharing raw data. This strategy has been advocated in education (Hridi et al., 2024) to preserve privacy while enhancing personalization. Together, these techniques allow Parrot to refine instructional support across deployments while complying with strict data protection standards.

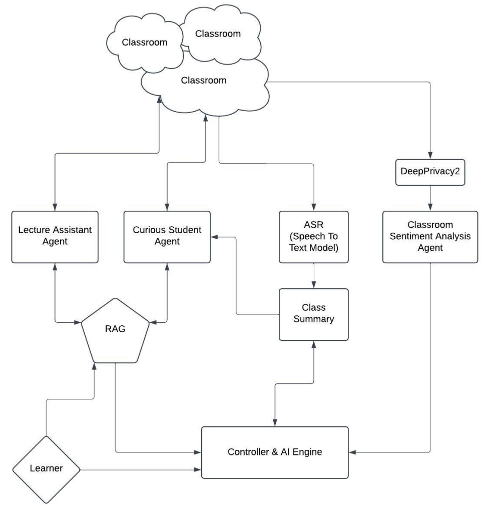

Figure 1. Architecture of the Classroom Agent System. The system includes several AI components to support classroom interaction. The Lecture Assistant Agent responds to course-related questions. The ASR (Automatic Speech Recognition) + Class Summary module captures spoken content from the classroom and periodically generates summaries. The Curious Student Agent uses these summaries to generate engaging, student-like questions. The Classroom Sentiment Analysis Agent, supported by DeepPrivacy2 for privacy protection, monitors visual and auditory classroom cues to classify student engagement (e.g., boring vs. interesting).

3. System Architecture

Parrot consists of modular components orchestrated into a real-time cooperative workflow:

Sensing: Audio and video feeds are processed in real time. DeepPrivacy2 anonymizes faces to preserve emotional cues like gaze or posture while protecting identity. Whisper-based ASR transcribes lecture speech.

Reasoning: Multimodal sentiment analysis combines visual, audio, and text cues to produce interpretable engagement scores. GPT-4o analyzes snapshots to infer mood and confusion.

Acting: Depending on engagement and topic complexity, Parrot either summarizes (Assistant Lecturer) or generates questions (Curious Student). Teachers preview outputs through a dashboard.

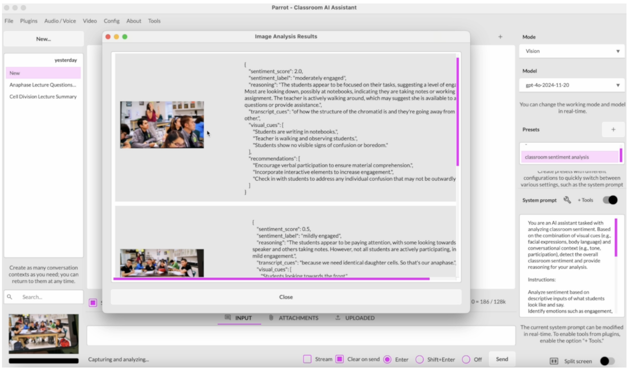

Figure 2. Classroom Agent Application Interface. The interface displays real-time analysis of classroom interactions between the lecturer and students. It visualizes sentiment analysis results derived from facial expressions, body language, and contextual cues, providing insights such as engagement levels, attention, and participation. The system also generates actionable recommendations to improve classroom dynamics.

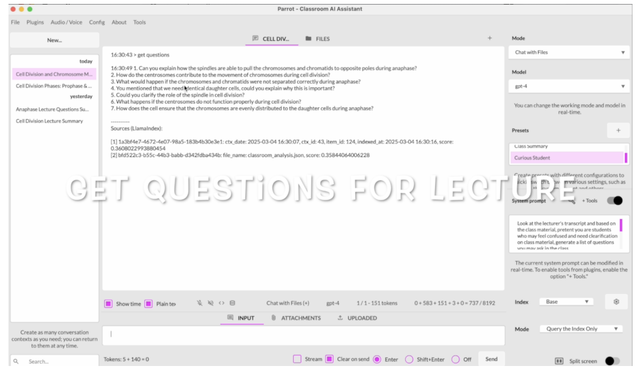

Figure 3. Curious Student Agent Interface. This module simulates an inquisitive student by analyzing the lecture transcript in real time and generating context-aware follow-up questions. It identifies moments that may require clarification and formulates thoughtful queries to promote deeper understanding and student engagement.

All agents operate semi-autonomously under a central Controller that manages timing and inter-agent communication. Retrieval-Augmented Generation (RAG) grounds questions and answers in syllabus content, with interpretable citations enabling teacher oversight.

The architecture (Fig. 1) follows a pipeline that processes classroom data into real-time insights and interventions. Key components include:

3.1. Multimodal Data Acquisition

Classroom Inputs: Parrot captures audio-visual input via a classroom camera and microphone. DeepPrivacy2 anonymizes faces while preserving expressions and gaze for sentiment analysis. High-accuracy ASR (e.g., Whisper) transcribes speech in real time, feeding language modules with clean input.

Speech-to-Text (ASR): An automatic speech recognition model converts the instructor's spoken words (and any spoken student questions) into text transcripts in real time. We use a high-accuracy ASR (e.g., Whisper or a domain-tuned model) to ensure the subsequent language processing modules receive quality input. The ASR is crucial for creating the Class Summary and for understanding context when the agent formulates questions or answers.

3.2. Sentiment Analysis Module

As showed in figure 2. This module fuses visual (face and pose), audio (tone), and textual (transcript) features into a sentiment score (e.g., from -5 to +5). Anonymized video ensures privacy, while GPT-4o supplements interpretation of class snapshots. Teachers view sentiment trends over time to identify emotional hotspots and adjust instruction accordingly.

3.3. Curious Student Agent

As showed in figure 3. Triggered during lulls or complex content, this module prompts GPT-4o to generate relevant, curiosity-driven questions based on lecture transcripts and summaries. Grounded in curriculum-aligned content, the questions are previewed by teachers for approval to avoid disruption and support inquiry-based learning.

3.4. Assistant Lecturer Agent

This module answers questions or provides clarifications, using RAG to retrieve syllabus-aligned content before prompting GPT-4o. Answers cite sources (e.g., textbook page or past lecture), boosting transparency and instructional trust. A short-term memory of lecture context helps maintain continuity.

3.5. Class Summary Generator

GPT-4 generates summaries periodically from the accumulating transcript and optionally slide data. Summaries highlight key points and include sentiment flags (e.g., confusion), serving as feedback for both teachers and students.

3.6. Curriculum-Grounded Retrieval via RAG

Rather than using a standalone knowledge graph, Parrot retrieves relevant syllabus-based content using semantic embeddings. This supports grounding for both Assistant Lecturer and Curious Student agents, enhancing content fidelity and minimizing hallucinations.

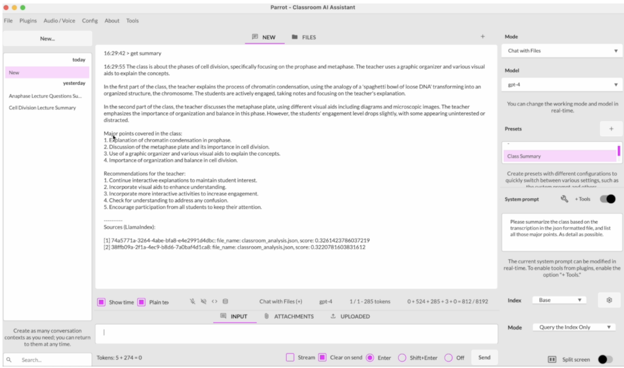

Figure 4. Lecturer Assistant Agent Interface. This component periodically analyzes lecture transcripts and generates structured summaries of key points covered in class. It also provides actionable recommendations for the teacher to maintain student engagement and improve instructional clarity.

Figure 5. Classroom Review Control Panel: visualization tracks student sentiment over time based on classroom interactions. The control panel enables lecturers to review class progress, identify moments of confusion or disengagement, and receive actionable suggestions—such as adding visual aids or promoting peer discussion—to enhance teaching effectiveness.

3.7. Classroom Review Control Panel

The Classroom Review Control Panel, illustrated in Figure 5, provides lecturers with an interpretable dashboard that aggregates real-time multimodal data, including student sentiment trends and lecture summaries. It visually highlights shifts in engagement, identifies specific cues (e.g., yawning, participation drop-offs) and offers context-aware, actionable suggestions, enabling lecturers to dynamically adjust classroom instruction. This facilitates immediate pedagogical adjustments, improving classroom responsiveness and educational effectiveness.

3.8. AI Engine (LLM) and Controller

At the heart of Parrot is the large language model (GPT-4o), which powers natural language understanding and generation for most of the modules. A central *Controller* orchestrates the pipeline: it manages timing (when to summarize, when to ask a question), directs inputs to the LLM with appropriate prompts, and merges outputs from different modules. For efficiency, many tasks run in parallel threads – e.g., ASR and sentiment analysis run continuously, while the LLM is invoked in bursts for summary or Q&A. We also implement a fallback: if the LLM is unavailable or too slow, the system can default to simpler behaviors (e.g., only do ASR transcription and hold questions for later), ensuring robustness. The LLM is the most computation-heavy component, so we optimize prompt sizes (by summarizing context) and use caching for repeated queries. In essence, Parrot's system architecture balances advanced AI capabilities with practical constraints of a classroom: it processes rich multimodal data, integrates domain knowledge, and interacts in human-like ways, all while providing outputs that a teacher can validate and use.

3.9. Learner: Adaptive and Federated Refinement

At the core of Parrot's agentic intelligence is the **Learner**—a subsystem responsible for improving classroom performance over time through adaptive feedback. All major modules—including Sentiment Analysis, Curious Student, Assistant Lecturer, and Class Summary—feed outcome signals into the Learner, which orchestrates both local adaptation and federated knowledge sharing. This enables Parrot to evolve from a reactive assistant into a proactive and context-sensitive collaborator.

Local Adaptation: Each Parrot instance tracks how students and teachers respond to its actions—such as whether engagement improves after a clarification, or which types of questions prompt discussion. Based on this feedback, it refines local behavior policies without changing the underlying LLM. For example, if open-ended questions consistently increase engagement in STEM classes, the Curious Student module will prefer those in future sessions. Likewise, if sentiment flags are overly sensitive in a particular classroom, the Sentiment Agent will adjust thresholds to match local norms.

Federated Collaboration: Rather than uploading raw classroom data or fine-tuning the LLM, Parrot shares anonymized usage metadata—e.g., which prompt templates correlated with better engagement, or which retrieved passages led to accurate answers. These lightweight updates are periodically aggregated across deployments to update a global prompt strategy library and shared retrieval corpus. This allows the global system to reflect diverse classroom conditions while preserving local privacy and complying with regulations like FERPA.

RAG and Prompt Policy Co-Evolution: Adaptation in Parrot targets two key levers: retrieval quality and prompting effectiveness. When certain documents consistently improve answer quality or comprehension, they are prioritized in the shared RAG index. Simultaneously, the Controller logs which prompt styles (e.g., Socratic, explanatory, scaffolded) lead to positive sentiment shifts, and these successful templates are surfaced for reuse. Over time, this creates a distributed memory of pedagogical best practices, refined through classroom use rather than offline engineering.

Together, these mechanisms make the Learner a central component of Parrot's agentic loop—sensing feedback, adapting actions, and sharing discoveries. This continual refinement supports trustworthy autonomy, contextual sensitivity, and scalable improvement across classrooms.

4. Experiments and Evaluation

We conducted a series of experiments to evaluate Parrot's performance across its core functionalities: sentiment analysis, lecture summarization, privacy protection and interactive Q&A. The evaluation used a combination of simulated classroom sessions and benchmark comparisons to thoroughly test the system under diverse scenarios. Below we outline the experimental setup, key evaluation metrics, and results for each module.

4.1. Experimental Setup

Data Collection: We compiled 15 hours of high school and college classroom video across subjects (science, math, literature), segmented into ∼3-minute clips. Each includes lecture and visible student reactions. Supplementary materials (slides, textbook excerpts, forums) populated the RAG knowledge base. A smaller set of annotated transcripts supported validation.

Sentiment Ground Truth: 100 video clips were annotated by humans for student engagement (scale: 5 to -5) at 20-second intervals, with affective events (e.g., laughter, confusion) flagged to provide labeled data for sentiment evaluation.

Baselines and Metrics: We compared Parrot to (1) *Text-Only Sentiment*, (2) *No-RAG LLM*, and (3) *No-Questions Mode*. Metrics included accuracy, precision, recall, F1 (especially for disengagement), question relevance, answer correctness, system latency, and safety failures.

4.2. Results: Sentiment Analysis

Parrot achieved strong performance in detecting student engagement, outperforming unimodal baselines (see Table 1). It reached an overall accuracy of **0.92** and an F1-score of **0.90** for disengagement detection, spotting subtle cues like distraction ahead of human observers. The text-only baseline underperformed on recall, highlighting the benefit of multimodal fusion. High-engagement events (e.g., laughter) were also identified with **0.95** precision.

Table 1. Comparison of Sentiment Analysis Techniques

Technique	Precision	Recall	F1-score
Text-only	0.78	0.72	0.75
Visual-only	0.84	0.86	0.85
Parrot	**0.91**	**0.89**	**0.90**

Table 2. Sentiment Performance Before and After Anonymization

Type	Precision	Recall	F1-score
Original	0.91	0.89	0.90
Anonymized	0.90	0.88	0.89

4.3. Results: Privacy-Preserving Sentiment with DeepPrivacy2

Using DeepPrivacy2-anonymized video, Parrot retained nearly identical performance. Engagement classification remained high (**F1 = 0.89**) with minimal loss compared to raw video (see Table 2). This confirms that privacy-preserving processing can coexist with accurate sentiment detection.

4.4. Results: Speech Transcription and Summarization

Whisper ASR reliably transcribed lectures with minor errors (e.g., misnamed terms), enabling effective downstream tasks. The summarizer generated clear, sentiment-aware digests, averaging a usefulness rating of **4.7/5**. Compared to a sentiment-agnostic baseline, Parrot's summaries were seen as more insightful—often highlighting confusing topics or forgotten details.

4.5. Results: Interactive Q&A (Curious Student and Assistant Lecturer)

Parrot answered **94%** of instructor-validated questions correctly, dropping to **80%** without RAG, which shows the value of curriculum grounding. The Curious Student agent asked 3–8 questions per session, with **17/20** rated highly relevant. Several prompts helped re-engage classrooms during lulls, confirming its role in attention recovery and inquiry stimulation.

4.6. Efficiency and Real-Time Performance

Parrot ran in real time on a single-GPU machine. Sentiment updated every 20s, ASR lagged 1–2s, and LLM outputs appeared within 2–5s. Strategic frame sampling (1 per 10s) kept latency low. This confirms feasibility for real-time classroom integration with minimal delays.

4.7. Adaptive Learning and Federated Collaboration

The Learner module enables Parrot to adapt over time without modifying core models. Locally, it refines question styles and sentiment thresholds based on classroom response (e.g., favoring open-ended questions in STEM lectures). Globally, it contributes anonymized metadata—such as successful prompts or effective retrieval passages—to a shared pool used to update the prompt library and RAG index.

Local Adaptation: In simulated classroom runs, the Learner tracked engagement outcomes and adjusted agent behavior. For example, after observing higher sentiment scores in response to exploratory prompts, the Curious Student module shifted toward open-ended question templates—rising from 43% to 67% usage across three STEM lectures.

Federated Collaboration: In a simulated federated setup with three agents, the Learner module contributed updates that improved question generation and answer quality across agents. After aggregation, the relevance of AI-generated questions increased by 8%, and retrieval alignment with syllabus materials rose from 85% to 93%.

This federated approach protects student data while allowing distributed improvement. Across deployments, Parrot evolves by learning which interventions work best in diverse settings. These findings validate the system's capacity for continuous learning and contextual sensitivity without compromising privacy.

5. Discussion: Toward Agentic, Cooperative AI in Education

Parrot is not a tool but an agent—it acts independently, collaborates adaptively, and communicates transparently. Its success highlights several core principles of agentic intelligence:

- **Autonomy**: Parrot acts without scripting. Its question generation, summarization, and sentiment responses are data-driven and context-sensitive.
- **Collaboration**: All actions are presented as cooperative aids, not replacements for teacher decisions. The agent supports human judgment, never overrides it.
- **Interpretability**: Rationale and citation mechanisms provide traceability. Teachers retain full agency via oversight interfaces.
- **Adaptivity**: Parrot continuously adapts its models based on classroom-specific feedback. It adjusts its question strategies and sentiment thresholds in response to engagement metrics, ensuring context-aware personalization over time.
- **Federated Collaboration**: All Parrot agents participate in a federated learning system focused not only on model parameter tuning but also on knowledge and reasoning policy enhancement. Specifically, each classroom instance monitors the effectiveness of its Retrieval-Augmented Generation (RAG) responses—logging which document passages best supported student learning—and contributes high-yield excerpts to a shared retrieval index. Simultaneously, the Learner module refines its prompt templates based on local interaction outcomes (e.g., which engagement-triggering prompts led to participation spikes). These anonymized updates are aggregated to construct a global RAG corpus and adaptive prompt library, improving instructional logic across deployments while preserving data privacy.*These refinements are coordinated by the Learner module, which serves as the system's adaptive core—closing the feedback loop between sensing, acting, and improving.* This federated collaboration strengthens the system's contextual awareness and pedagogical sensitivity over time.
- **Privacy-by-Design**: Sensitive visual data is anonymized in real time; only derived engagement signals are retained.

These qualities mark Parrot as an example of agentic intelligence fit for complex, high-trust environments like education. Future work includes expanding to multi-agent coordination (e.g., group learning agents), more granular personalization models, and deeper integration with classroom analytics platforms.

6. Conclusion

Parrot represents a step toward agentic, interpretable, and cooperative AI in education. As an autonomous classroom agent, it observes, interprets, and contributes to real-time instruction while maintaining privacy and transparency. Its multimodal reasoning and collaborative outputs support—not supplant—human educators. Through its architecture and performance, Parrot demonstrates what it means for an AI system to act as an agent, not just a tool: it senses, adapts, explains, and partners. We invite further discussion on how such agentic intelligence can transform human-AI workflows across educational and other complex domains.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Brusilovsky, P. and Millán, E. Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology*, 38(4):618–628, 2007. doi: 10.1111/j.1467-8535.2006.00658.x.

D'Mello, S. K. and Kory, J. A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, 47(3):43:1–43:36, 2015.

Graesser, A. C., Chipman, P., Haynes, B. C., and Olney, A. Autotutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48 (4):612–618, 2005.

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., and Aleven, V. The classroom as a dashboard: Co-designing wearable cognitive augmentation for k-12 teachers. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK)*, pp. 79–88, 2018.

Hridi, M., Rana, M. M., Shokri, R., and Dillenbourg, P. Federated learning in education: A survey on opportunities, privacy, and real-world implementation challenges. *IEEE Transactions on Learning Technologies*, 2024. URL https://arxiv.org/abs/2403.06597. To appear.

Hukkelås, H., Mester, R., and Lindseth, F. Deepprivacy2: Real-time facial anonymization using generative adversarial networks. *arXiv preprint arXiv:2310.11322*, 2023.

Khan Academy. Khanmigo: Ai-powered tutoring with gpt-4. https://www.khanacademy.org/khan-labs, 2023. https://www.khanacademy.org/khan-labs.

Khosravi, H., Kitto, K., Lopez, T., and Gašević, D. Explainable artificial intelligence in education: A systematic review. *Computers & Education: Artificial Intelligence*, 3:100082, 2022.

Lee, J., Kim, S., and Tan, J. Vidaas: Vision-based automated assessment system for classrooms. In *Proceedings of the 2024 ACM CHI Conference on Human Factors in Computing Systems*, 2024.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kulkarni, D. P., Levy, O., Lewis, M., Yih, W.-t., et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Long, A., Chen, M., and Zhou, H. Gpt-4o as a classroom analyst: Automatic coding and feedback on dialogue. In *Proceedings of the 2024 International Conference on Artificial Intelligence in Education (AIED)*, 2024.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1273–1282. PMLR, 2017.

Pan, X., Zhang, Y., Liu, K., and Wu, J. Multimodal biometric emotion detection in classrooms. In *Proceedings of the 2024 International Conference on Learning Analytics & Knowledge (LAK)*, 2024.

U.S. Department of Education. Family educational rights and privacy act (ferpa). https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html, 2023. Accessed: 2025-05-07.

U.S. Department of Education, Office of Educational Technology. Artificial intelligence and the future of teaching and learning: Insights and recommendations. Technical report, U.S. Department of Education, 2023. https://tech.ed.gov/ai.

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., and Movellan, J. R. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 5(1):86–98, 2014.