Glean: Capturing Knowledge Beyond Slides

香港教育大學
The Education University of Hong Kong

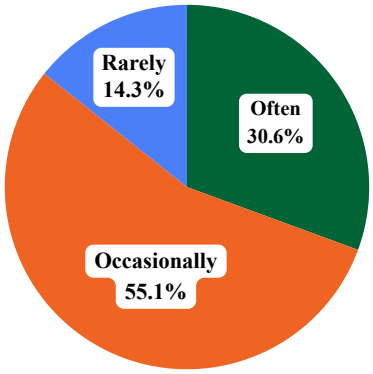
Jiaqi Gao, Jiaqi Huo, Tattibekov Assylzhan, Cheuk Him Chan, Yihan Deng, Wenjun Yuan

Motivation

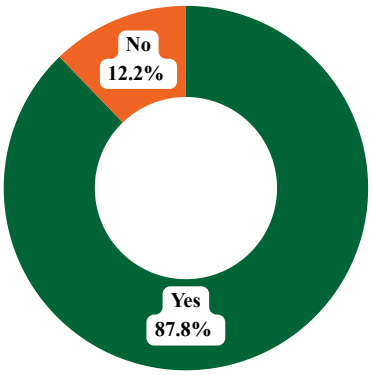
Traditional lectures often suffer from the loss of "unwritten brilliance"—the impromptu explanations, real-world examples, and dynamic interactions that disappear once the class ends. This disconnect between static courseware and dynamic teaching creates a critical gap in knowledge retention, leaving students with incomplete notes that fail to capture the full depth of instructional intent. Our research addresses this gap through Glean, a system designed to preserve and structure the knowledge beyond slides.

Survey & Data Analysis

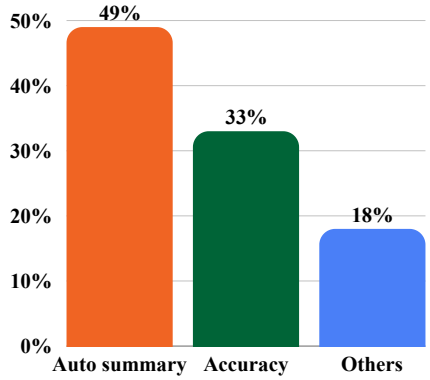
Missing Insights



Impact on Learning



User Priorities



Key Findings

- Our survey encompassed diverse stakeholders—teachers, parents, and students.
- 87.75% report missing impromptu insights significantly affect their learning.
- 85.71% expect an automated solution to capture non-textual knowledge.
- 53.06% prioritize local data storage, indicating privacy concerns.

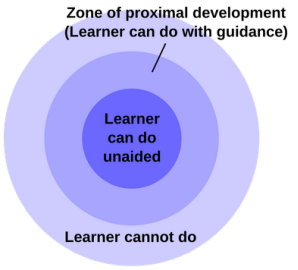
Problem Statement

> How can we enhance **knowledge transfer** by empowering teachers and students to **capture and structure** the impromptu, non-textual teaching beyond **limited slides**?

Research Objectives

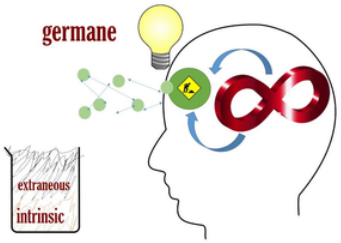
- **Capture & Preserve:** Extract and preserve teacher's impromptu explanations and real-time insights that occur beyond prepared slides.
- **Structure & Anchor:** Automatically synchronize audio with slide content to create permanent, searchable study resources.
- **Enhance Engagement:** Reduce cognitive load during lectures, enabling students to focus on active listening rather than frantic note-taking.
- **Support Learning:** Provide accessible revision materials that support long-term knowledge retention and deeper understanding.

Education Theory Base

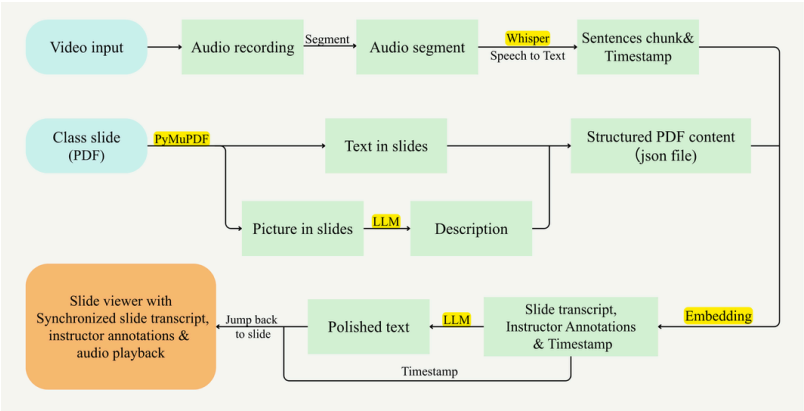


Cognitive Load & Transient Info: Manual note-taking of transient verbal information creates high extraneous cognitive load, splitting attention between listening and transcription (Wong et al., 2012). This cognitive overload impedes deep processing and retention. Glean reduces extraneous load by automatically capturing audio via Whisper STT, synchronising with slides, and using AI to structure notes—freeing students' cognitive resources for germane learning tasks and active engagement.

Constructivism & ZPD: As a constructivist scaffold (Piaget, 1954), Glean transforms ephemeral impromptu teaching into permanent, searchable learning resources. By anchoring teacher insights onto slides, it supports Vygotsky's Zone of Proximal Development (Cole & Scribner, 1978), enabling students to internalize and construct knowledge through interaction with structured, accessible materials rather than struggling with incomplete notes.



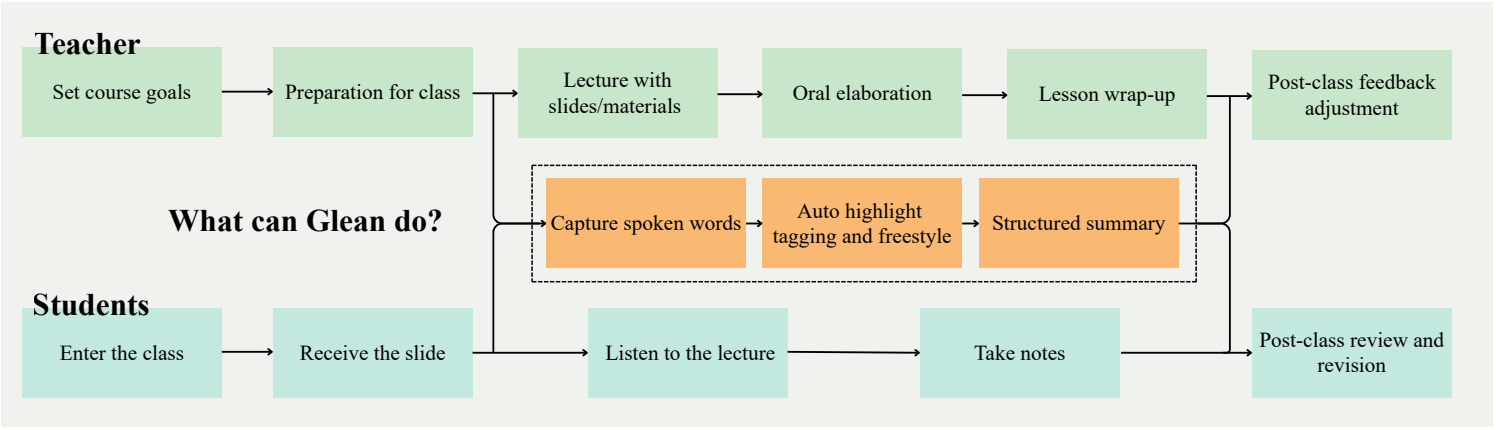
Implementation



Why we chose these methods:

- Audio & Whisper (STT): Accurately captures spoken explanations from lecture recordings, ensuring no critical verbal insights are missed.
- Vector Embedding: Contextually synchronizes the transcribed audio with the corresponding text in the presentation slides.
- Freestyle & AI Optimization: Utilizes an AI model to refine, summarize, and structure the combined multi-modal data into coherent, easy-to-read study notes for students.

AI Augments Humans: Scenarios & Impacts



Tolerable Scenario: AI augments not replaces human teaching and learning. LLMs have **hallucinations** and **black-box** issues, making direct educational use dangerous. Our project avoids full AI generation. Instead, we use embedding-based text comparison for accurate concept matching and verification. LLMs are limited to text polishing only—enhancing clarity without changing meaning. This preserves classroom authenticity, supports teachers, and encourages student engagement while eliminating AI misinformation risks.

Reference

Cole, M., & Scribner, S. (Eds.). (1978). Mind in society: The Development of Higher Psychological Processes. *Harvard University Press.*

Piaget, J. (1954). The construction of reality in the child. Routledge & Kegan Paul.

Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction*, *22*(6), 449-457.