

From Course Concept to Lecture Video: An AI-Powered System for Automated MOOC Development

Igo, J., Yao, H., Xu, W., Kellam, N., & Wei, H.

AI-generated video

ICAP framework

MOOC development

Producing lecture videos for online courses requires significant time and expertise, and current AI-generated videos often lack accuracy, consistency, and pedagogical grounding. This study presents a controllable system that converts slides and scripts into lecture videos. The system follows instructor intent and integrates the ICAP framework to ensure the presence of constructive questions. We evaluate the system using four metrics: ICAP assurance, video generation time ratio, speech consistency, and slides coverage rate. Results show that the system provides higher controllability and stronger educational alignment than NotebookLM, demonstrating its potential for scalable MOOC development.

Introduction

Massive Open Online Courses continue to expand access to higher education, yet instructors face three major challenges when producing lecture videos. First, video creation is highly time-consuming and requires substantial effort in design and production, which slows course development and affects the quality of online materials (Hollands and Tirthali, 2014; Guo et al., 2014). Second, current video generation AI tools like NotebookLM (Google, 2025) often produce inconsistent or inaccurate content and cannot reliably follow instructor-provided slides or scripts. Lastly, existing systems do not incorporate learning science principles, which reduces opportunities for meaningful cognitive engagement (Bauer, 2025).

This study presents an automated system that converts course slides and scripts into lecture videos through a controllable generation process to address these challenges. The system ensures accurate alignment with instructor intent, maintains full coverage of the instructional materials, and integrates the ICAP framework (Chi and Wylie, 2014), which characterizes different levels of cognitive engagement: Interactive, Constructive, Active, and Passive. By embedding constructive and interactive questions, the system encourages learners to move beyond passive reception and engage in deeper processing of instructional content.

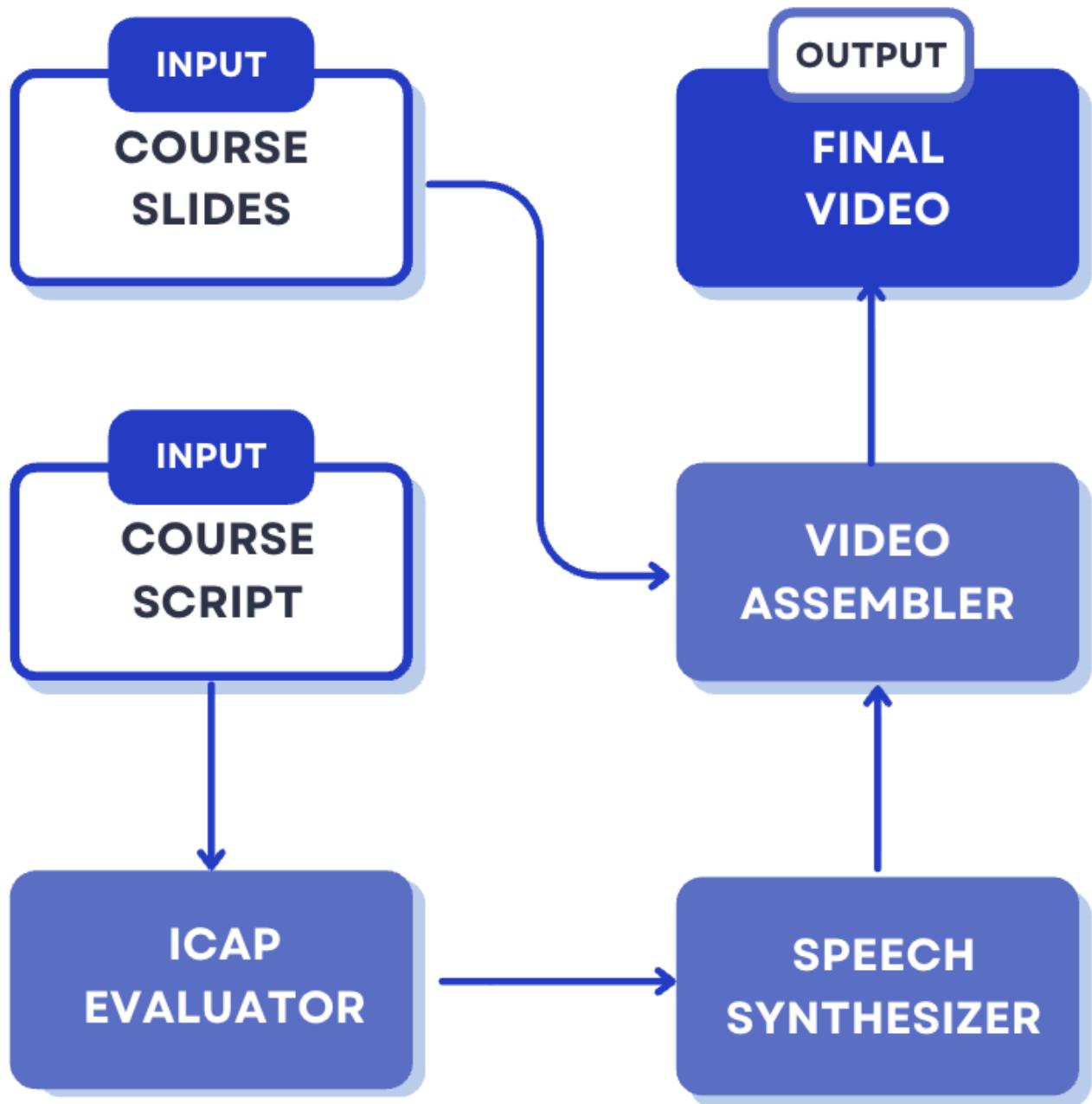
Our contributions are threefold. First, we propose a controllable pipeline for automated MOOC video generation. Second, we integrate ICAP-based design principles to promote higher levels of learner engagement. Finally, we evaluate the system using metrics that directly address current challenges in consistency, efficiency, and pedagogical quality.

Methodology

The proposed system automatically generates MOOC lecture videos by transforming course materials into coherent presentations. It comprises three sequential modules ICAP Evaluator, Speech Synthesizer, and Video Assembler that ensure pedagogical alignment, generation efficiency, and content consistency under the ICAP framework, as shown in Figure 1.

Figure 1.

Overview of the proposed pipeline. Course slides and scripts are used as inputs, with the script evaluated by the ICAP module before speech synthesis and video assembly produce the final lecture video.



The system takes course slides and accompanying scripts as input. Instructors may upload pre-existing materials, or, when only a course title or outline is provided, Instructional Agents (Yao et al., 2025) automatically generate slides and scripts aligned with the specified learning objectives using large language models. Subsequently, the ICAP Evaluator employs a large language model to analyze the slides and refine the lecture scripts in accordance with the ICAP framework, embedding guiding questions and constructive learning elements to foster cognitive engagement and higher-order thinking, while maintaining content consistency. The optimized instructional script is processed through a Speech Synthesizer, specifically, Google Text-to-Speech, to generate the lecture narration. The synthesizer modulates speech rate, intonation, and volume to ensure clarity and listening comfort. The resulting audio is then validated against the optimized script to guarantee precise alignment between text and narration, thereby eliminating discrepancies between spoken content and visual materials. In the final stage, the Video

Assembler synchronizes slide frames with corresponding audio segments and compiles them into a complete lecture video using the FFmpeg framework, producing a cohesive and pedagogically aligned output.

Experiment

Evaluation Setting

We employ GPT-4o-mini as the base model for the ICAP Evaluator. The system was evaluated using four primary metrics: (1) ICAP Assurance checks whether each video includes constructive questions or activities that align with the constructive level of the ICAP framework, implemented as interaction-eliciting instructional cues intended to encourage learners' active reasoning, reflection, or self-explanation. (2) Generation Time Ratio is the total video generation time divided by the duration of the generated video. A lower value indicates faster generation. (3) Speech Consistency evaluates the match between generated speech and input script based on ROUGE-1 score (Lin, 2004). A higher value indicates a better consistency. (4) Slides Coverage is the ratio of input slides that appear in the final output video. The evaluation was conducted on five university-level courses comprising hundreds of slides, with Google NotebookLM used as the baseline.

Main Results

As shown in Table 1, the proposed system outperforms Google NotebookLM across all four metrics. Our system has a lower Generation Time Ratio, addressing the time-consuming nature of lecture production. Our system also has perfect Speech Consistency and Slides Coverage scores, indicating precise alignment with instructor scripts and complete use of provided materials, overcoming the issue of inconsistent or inaccurate content. Finally, ICAP Assurance is guaranteed in our system, ensuring the intentional integration of constructive learning prompts and addressing the lack of learning-science principles in existing tools.

Table 1.

Results of our system compared with NotebookLM. For Generation Time Ratio, the lower the better; for Speech Consistency and Slides Coverage Rate, the higher the better.

Method	Generation Time Ratio (↓)	Speech Consistency (ROUGE-1, ↑)	Slides Coverage Rate (% , ↑)	ICAP Assurance*
Ours	0.29	100.00	100.00	Guaranteed
NotebookLM	0.91	19.82	39.81	Incidental Only

*ICAP Assurance refers to the intentional and guaranteed presence of constructive prompts. NotebookLM may produce constructive-style guiding questions, but these occur incidentally and are not controlled by an instructional design framework.

Discussion

The results show that controllability and instructional alignment improve the practicality of AI-generated lecture videos. The system consistently satisfied ICAP Assurance, produced accurate narration, preserved full slide coverage, and generated complete videos within a predictable time. These features address common issues in existing AI video tools and allow instructors to create pedagogically aligned materials with reduced effort. This supports more scalable MOOC development.

Future work will integrate a Learning Collaborator Agent to introduce interactive learning opportunities within the ICAP framework.

References

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., and Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review*, 37(2), 45.

Chi, M. T., and Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243.

Google, 2025. NotebookLM [Computer software]. <https://notebooklm.google/>

Guo, P. J., Kim, J., and Rubin, R. (2014, March). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the First ACM Conference on Learning at Scale* (pp. 41-50).

Hollands, F. M., and Tirthali, D. (2014). MOOCs: Expectations and Reality. Full Report. Online Submission.

Lin, C.-Y. (2004, July). ROUGE: A package for automatic evaluation of summaries. *Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL*, Barcelona, Spain.

Yao, H., Xu, W., Turnau, J., Kellam, N., and Wei, H. (2025). Instructional agents: LLM agents on automated course material generation for teaching faculties. *arXiv preprint arXiv:2508.19611*.

