

PDF2Lec: Streamlining LLM-driven Lecture Generation

Rachel Lin¹, Mutian Hong¹, Chuyan Zhou¹, Jan Chen¹, and Lucy Struefing¹

¹UC Berkeley, Team: PDF2Lec, Application Track, 3 Units

Abstract

To bring a next-gen learning experience through AI-driven teaching, we are developing a lecture audio creator system called PDF2Lec. Addressing the lack of interactivity in prior solutions, our application allows users to customize lecture generation and incorporates a chatbot for a personalized educational experience. The process begins with a guard agent verifying uploaded PDF lecture slides. Once verified, similar slides are merged, and a slide parsing agent, powered by the OpenAI GPT-4o model, evaluates slide complexity and uses keywords to perform retrieval-augmented generation (RAG) from a textbook, if provided. The gathered information is then passed to the lecture generating agent, which produces lecture text that will later on be converted into audio using OpenAI’s Text-to-Speech (TTS) model. We conducted user testing by having participants watch both our AI-generated lecture and the original lecture. Participants were assessed through quizzes to measure comprehension, and feedback was gathered on lecture quality. Results revealed that users who listened to the AI-generated lecture performed 11.21% better on average. Furthermore, the AI-generated lectures received higher ratings in clarity, depth, and engagement. This preliminary analysis serves as a benchmark and highlights a promising approach to enhancing learning experiences using LLMs. For future work, our goal is to incorporate additional interactive features, such as dynamic annotations, to create more effective presentation for concept-heavy topics.

1 Introduction

Much work has been dedicated to generating audio content from textual sources. One such example is the generation of conversational podcast scripts from research paper PDFs using large language models (LLMs) in academia (1). Similarly, LLMs are revolutionizing healthcare by converting medical papers into auditory plastic surgery education (2). These advancements enhance accessibility for diverse learning styles and enable auditory learners to engage more effectively with complex information across various fields.

State-of-the-art practices often involve leveraging LLMs to extract key textual elements from source material and then generating text scripts based on the information gathered. These scripts are then converted into audio using Text-to-Speech (TTS) technology, which employs natural language processing to produce high-quality, natural-sounding voices (3). However, despite these advancements, current AI-generated educational content lacks personalization, interactivity, and adaptability. While previous studies have focused on creating video content from pre-written speaker notes to make education more accessible, they have not addressed the challenge of generating content interactively with students. To address these limitations, we propose a novel system for generating segmented audio lectures directly from PDF slides without requiring pre-written speaker notes. This system uses advanced AI techniques to analyze slide content, employing retrieval-augmented generation (RAG), and generate coherent audio segments. It goes beyond static delivery by incorporating a "human-in-the-loop" approach (4), enabling learners to ask questions and receive real-time clarifications through a chatbot.

Overall, our contributions are as follows:

1. Proposes a potential application to enhance interactivity with students, leveraging AI to create a more engaging experience through customization for different learning styles and real-time clarification through a chatbot.
2. Evaluates the educational capabilities of multimodal LLMs in low-resource scenarios, providing insights into their effectiveness and limitations compared to human teachers.
3. Tests instructional prompts and provides examples that yield effective teaching outcomes.

2 Related Works

2.1 Large Language Models in Generated Educational Content

Large Language Models have advanced significantly in generating complex outputs, including educational applications. As Denny *et al.*, [5] highlights, LLMs can enhance learning by automatically creating diverse educational resources across domains and skill levels. Building on this, our project integrates LLMs to produce personalized, interactive lecture audio, prioritizing real-time interactivity and customization over static resource generation.

2.2 Preceding works about Lecture Content Generation

AI-generated educational content has advanced significantly, with applications like automated video lectures and AI-powered MOOCs. Dao *et al.*, [6]’s work, titled *AI-Powered MOOCs: Video Lecture Generation*, utilizes technologies like Text-to-Speech and Speech-driven-Face (SDF) to automate video production, integrating an instructor’s voice and face without manual recordings. Their work, like ours, seeks to enhance study efficiency, reduce instructor workload, and improve accessibility. However, while their approach focuses on the visual aspects of lecture content, our work emphasizes audio-based lecture generation. Using LLMs and RAG, we create dynamic, interactive audio content, prioritizing detailed and adaptable material over visual elements like an instructor’s face or gestures.

2.3 Audio Generation

Another noteworthy project in AI-driven audio generation is Podcastfy by Souza [7], which converts textual content into podcast-style audio using advanced text-to-speech technologies. It repurposes static written content like blogs and articles in conversational, customizable formats. This versatility makes it applicable to both entertainment and education. While Podcastfy focuses on engaging, conversational podcasts, our work adapts this concept to education by producing structured lecture audio. Unlike Podcastfy’s static outputs, our system uses RAG and a chatbot to enable real-time learner engagement and personalized content delivery.

Furthermore, NotebookLM by Google [8], a notetaking framework, also includes a podcast feature that provides audio overviews of text. However, a notable limitation of NotebookLM is its lack of accompanying transcripts or synchronized visual aids. This can pose challenges for learners who benefit from visual reinforcement or prefer to follow along with text. To address these limitations, our project combines audio generation with synchronized lecture slides, offering a multimodal learning experience that combines auditory and visual components.

3 Methodology

The Pdf2Lec pipeline transforms PDF-based input educational materials (such as lecture notes, slides, or even textbook chapters) into a multimedia learning experience powered by LLMs. This pipeline produces customized lectures presented as videos with synchronized audio narration and text captions. Additionally, it supports interactive question-answering during playback, enabling a dynamic, personalized learning environment.

We discuss the detailed generation algorithm (process) below. From now on, we denote the agent as well as the model as θ , because what determines the property of the agents is predominantly the conditioned prompts p_* . We can also view θ as an abstract layer of different agents.

3.1 Preprocessing

3.1.1 Input and Initialization

The process begins with the user uploading a PDF file containing lecture materials. Optionally, the user may also provide supplementary resources such as textbook chapters or other textual

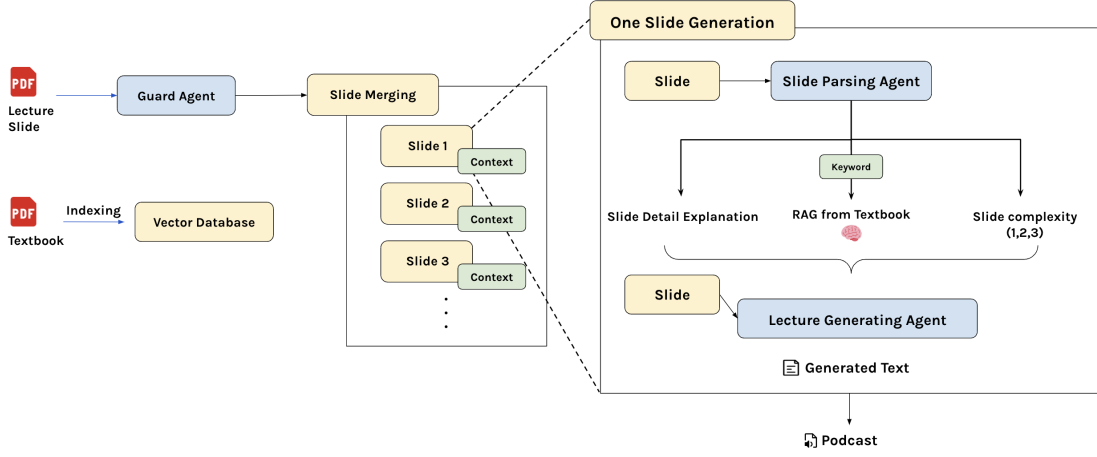


Figure 1: The workflow of uploaded PDF which a guard agent checks. Textbook upload is optional. Lecture slides are merged and for each slide we will generate text which will be converted to a podcast.

documents. Alongside the PDF upload, various settings can be configured, including teaching style and target complexity level.

3.1.2 Guard Agent Check

The uploaded PDF will be processed by the Guard Agent, which performs the following functionalities: **1. Validation of Input Material:** The Guard Agent will analyze selected pages of the slide to verify if the content aligns with the intended course material. If the input does not correspond to an appropriate course slide, the Guard Agent will raise an error and terminate the process. **2. Lecture Tagging:** Based on the content analysis, the Guard Agent will generate two tags for the lecture: A major tag representing the primary subject area (e.g., Computer Science). A subtitle tag representing the specific topic within the subject area (e.g., Machine Learning).

3.1.3 Slide Merging

We can denote the original PDF as a sequence of slides: $m_{1:T'}^0$. The merging algorithm which creates $m_{1:T} = \text{merge}(m_{1:T'}^0)$ leverages similarity thresholds to group related images and employs structured arrangements such as side-by-side layouts or grids for visibility. For example, if we group 4 slides together, the merged slide will have a 2×2 layout. We merge sequentially which is somehow trivial with a upper limit of the cardinality of the groups.

The merging process is designed to better accommodate complex instructional materials such as flowcharts and animations often found in lecture slides. By grouping and merging visually similar consecutive images, presents a more genuine and contextual view of dynamic or sequential elements for the subsequent generation pipeline.

3.2 Lecture Transcription (Text) Generation

3.2.1 Main Workflow

The core of the lecture transcription generation process lies within the primary loop. Each iteration of this loop involves analyzing a merged image as the input, extracting and organizing the necessary content by the slide parsing agent, and feeding this into the model for generating the content for the next page of slides by the generation agent. Also, extra parts will be placed before and after the generated content by the introduction & summary generation agent (the digesting agent) after the loop mentioned; and RAG as an option is also available during the text generation process.

3.2.2 Slide Parsing Agent

The Slide Parsing Agent is designed to interpret two primary inputs: the merged image representing the slide at a given time step and the context generated from the previous two slides. This context enhances coherence and continuity in the parsing process, leveraging the multi-modal capabilities of the OpenAI GPT-4o model.

The output of the Slide Parsing Agent is structured in a JSON format, encompassing three primary functionalities:

- 1. Detailed Slide Description:** The agent generates a textual description that captures the core visual and contextual elements of the slide. This description provides essential context for subsequent processing steps.
- 2. Keyword Extraction:** Specific keywords are extracted by analyzing the image. These keywords are used as retrieval keys in the RAG system.
- 3. Complexity Assessment:** The complexity level of the slide is categorized as either *brief*, *default*, or *detailed*. This metric informs the Lecture Generation Agent, dynamically adjusting the generated content to match the intended level of detail. The complexity level can either be determined automatically by the agent’s analysis of the slide or set to a fixed value by the user.

3.2.3 Page-by-page Generation

The system processes slides sequentially using GPT-4o to combine visual, contextual, and retrieved data. The inputs for generating content for a slide include:

- 1. Image:** The merged current slide.
- 2. Detailed Slide Description:** The structured content representation, extracted by the Slide Parsing Agent.
- 3. RAG Information:** If RAG is enabled, the model uses retrieved information generated from queries based on the extracted keywords.
- 4. Context:** A window containing the generated text from previous slides, ensuring coherence across the generated content. By default, this context includes content from the last two slides.
- 5. Prompt:** A well-designed prompt guiding the agent to generate engaging content. The output length and level of detail are dynamically adjusted based on the complexity level determined by the Slide Parsing Agent.

The combination of these inputs ensures that the generated content for each slide is coherent, relevant, and appropriately detailed.

3.2.4 Generation for the Introduction and Summary (Digest Agent)

Besides page-by-page generation introduced above, for a comprehensive and realistic experience of users watching the generated content, also inspired by previous work that generates lecture summaries by AI (9), we also stitch a paragraph of introduction placed before the whole lecture text (called the **introduction**, $\tilde{x}$), and after the text as well (called the **summary**, $\tilde{x}'$). The model used on this task is GPT-4o mini.

Here we model the generation of the introduction or summary as

$$\tilde{x} = \arg \max_x \mathbb{P}(x|x_{1:T}, p_{\text{intro}}; \theta), \quad \tilde{x}' = \arg \max_x \mathbb{P}(x|x_{1:T}, p_{\text{summ}}; \theta)$$

where p_{intro} is the prompt designed for generating the introduction and p_{summ} is for the summary. Note that θ here is for the digesting agent.

3.2.5 Multi-granularity Generation

The text generation process supports two approaches for determining the complexity level (or granularity) of the generated lecture content:

1. **User-Defined Complexity:** Users can manually select the desired level of complexity, choosing between options such as **brief**, **default**, or **detailed**.
2. **Automated Complexity Selection:** Alternatively, the system can utilize the results of the PDF parsing process to automatically determine the appropriate granularity for each section, which is handled by the slide parsing agent (3.2.2) we have mentioned above.

The granularity (complexity, g) choosing can be formalized as the process of generating p_{page} :

$$p_{\text{page}} = \begin{cases} \text{choose}(g), & \text{if user specifies,} \\ \text{choose}(\hat{g}(m_{1:T}; \theta)), & \text{otherwise,} \end{cases}$$

where θ is the slide parsing agent.

3.3 Audio Generation and Question Answering with RAG

3.3.1 Audio Generation from the Generated Text

The generated lecture text is converted into speech using OpenAI’s TTS model, producing the “lecture recording audio.” The system supports various models and voice types to provide diversity in audio content. Each audio segment corresponds to a specific section of the lecture text, enabling seamless integration into videos.

3.3.2 Question Answering and Retrieval-Augmented Generation

A simple QA agent allows users to ask questions during the lecture, enhancing interactivity. When a textbook is uploaded as supplementary input, the system employs RAG (10) to provide more accurate and contextually relevant responses.

The RAG mechanism is powered by FAISS (Facebook AI Similarity Search) (11), a high-performance library for embedding-based similarity search. FAISS enables efficient retrieval of relevant information from the textbook, addressing limitations of traditional query search methods. This grounding mechanism is utilized both during **lecture text generation** and **question answering**, ensuring that responses and generated content are enriched with precise, context-aware information. However, ensuring the relevance of retrieved content remains critical, as irrelevant information from the textbook can occasionally misalign outputs.

3.4 Deployment

The **PDF2Lec framework** is fully deployable as an application, with the source code available at ¹. The deployment integrates back-end and front-end systems, asynchronous task processing, metadata management, a user system, and database integration to support the functionalities described in Section 3. More details about our framework can be found in Appendix B.

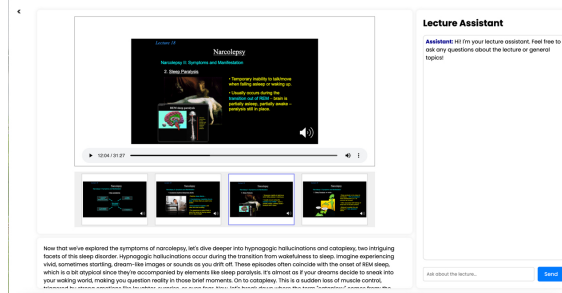


Figure 2: Front-end displaying generated content with transcript and chatbot.

¹https://github.com/itsuitsuki/pdf2lec_llm

4 Evaluation

We evaluate our product using two complementary approaches:

1. **Key Concept Recall:** Measuring the extent to which core concepts are covered in AI-generated lectures compared to baseline systems.
2. **User Testing:** Assessing user satisfaction, perceived lecture quality, and learning outcomes through controlled experiments with human participants.

4.1 Key Concept Recall

4.1.1 Experimental Setup and Evaluation

We selected six lecture slide decks, evenly distributed across STEM and non-STEM domains, for a balanced evaluation. Each slide deck consisted of an excerpt of around ten representative slides containing a mix of textual, graphical, and flowchart content. A list of key points was annotated manually for each slide and recorded in a reference sheet. These annotated key points served as the ground truth for evaluating the quality of the generated lectures.

Our system processed the original, non-annotated slide decks to produce lecture outputs. We tested multiple configurations of our PDF2Lec system with various granularities, including the **Brief Model**, **Default Model**, **Detailed Model**, and the **Auto Granularity Model with RAG**. We then compared our models to two baseline systems: **NotebookLM** and **GPT-4o (Zero-Shot)**, both of which were prompted with the same instruction: “*Create a detailed lecture-like presentation for each of these slides that covers everything.*”. These baselines provide a reference for the effectiveness of our tailored approach to lecture generation.

The evaluation process involved checking how many of the annotated key points from each slide were covered in the corresponding generated lecture. A score was assigned based on the proportion of key concepts accurately covered, with higher scores indicating better performance. Scores for all slides were averaged to determine the overall Key Concept Recall for each configuration.

4.1.2 Results

The results of the evaluation are summarized in Table I. Detailed information can be found in Appendix C.

Table I: Key Concept Recall Comparison Across Models and Configurations

Model Configuration	Social Science (%)	STEM (%)	Overall (%)
PDF2Lec (Brief)	89.34	84.47	86.90
PDF2Lec (Default)	96.74	92.58	94.66
PDF2Lec (Detailed)	97.82	95.64	96.73
NotebookLM	76.19	78.95	77.57
GPT-4o (Zero-Shot)	75.27	73.14	74.20
PDF2Lec (Auto) w/ RAG	99.21	97.73	98.47

Our Models: The results highlight distinct strengths and limitations of each configuration. **PDF2Lec (Brief)** performed the weakest, excelling at text-only slides or simple graphics but often providing vague or overly concise explanations for complex content like flowcharts or multi-layered graphics, leading to omissions of critical key points. **PDF2Lec (Default)** improved significantly, offering a balanced output suitable for most scenarios. It handled moderately complex slides well but occasionally struggled with highly intricate slides requiring deeper contextual understanding. **PDF2Lec (Detailed)** delivered the most comprehensive outputs, effectively capturing nuanced and complex concepts. However, its verbosity sometimes made simpler slides redundant or overwhelming for users. Finally, **PDF2Lec (Auto) w/ RAG** demonstrated the best performance by leveraging Retrieval-Augmented Generation for slides needing additional context, such as abstract or hypothetical problems. While effective, irrelevant textbook content occasionally distracted the model, highlighting the need for improved relevance filtering. Overall, the models show a clear

progression in handling complexity, with the RAG-enhanced configuration standing out for its ability to incorporate external context.

Baseline Models: NotebookLM and GPT4o (Zero-Shot): Both baseline models demonstrated reasonable performance on slides with straightforward textual content but struggled significantly with slides containing graphics or complex visual elements. While more sophisticated prompts or additional context could potentially improve their performance, these approaches still fall short of the tailored, multi-stage processing pipeline employed by our system, which explicitly processes visual content to extract meaningful context and achieve higher Key Concept Recall.

4.2 User Testing

4.2.1 Experimental Design

The experiment was conducted using two lecture topics: one STEM and one Social Science, both unfamiliar to the participants. Each topic included a single representative slide, for which we provided both an AI-generated lecture and the corresponding original recorded lecture. Participants were split into two groups:

- **Group A:** Participants first watched the original recorded lecture, answered a quiz based on the lecture, and evaluated its delivery. They then watched the AI-generated lecture and compared its quality to the original, providing satisfaction ratings.
- **Group B:** Participants first watched the AI-generated lecture, answered a quiz, and evaluated its delivery. They then watched the original recorded lecture and compared its quality to the AI-generated version.

4.2.2 Evaluation Metrics

The evaluation consisted of three sections:

Quiz Performance: Participants answered a quiz to measure their understanding of the lecture content. Quiz scores were used to compare the effectiveness of AI-generated lectures and original lectures in conveying information.

Lecture Quality Ratings: Participants scored the lectures on three criteria using a scale of 1 to 5: Clarity of Explanation, Depth of Understanding, and Engagement and Delivery.

Comparison Ratings: After watching both lectures, participants compared them, scoring satisfaction with the AI-generated lecture relative to the original.

4.2.3 Results

Quiz Performance: Participants who listened to the AI-generated lecture performed significantly better on quizzes compared to those who listened to the original lecture, with an average improvement of 11.21 percentage points. The improvement was especially pronounced in STEM topics, where AI-generated lectures led to a 15.15% higher quiz score, compared to a 7.27% increase for Social Science topics (Table II). More information can be found in Appendix D.

Table II: Quiz Performance Across Groups

Topic	Group A (Original)	Group B (AI Lecture)	Difference (B - A)
STEM	74.24%	89.39%	+15.15%
Social Science	89.39%	96.67%	+7.27%
Average	81.82%	93.03%	+11.21%

Lecture Delivery Ratings: Participants rated AI-generated lectures higher than original lectures across all dimensions: clarity, depth, and engagement. For the three criteria, on average, **AI-generated lectures received a score of 3.55 out of 5, compared to 2.49 for original lectures**, based on initial impressions (Figure 3). After participants viewed both lectures and had the opportunity to compare them, AI-generated lectures continued to score higher across all

criteria, further reinforcing their effectiveness and appeal (Figure 4). We acknowledge that the depth rating might have limited importance in this context because we purposefully selected a slide that led to a brief lecture and was relatively easy to understand. Real human lecturers often go deeper into topics, providing anecdotal insights or additional context beyond what is shown on the slides, which may offer a richer learning experience in more complex scenarios.

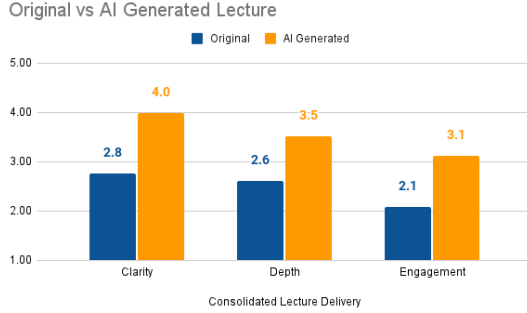


Figure 3: Lecture Delivery Ratings (Initial Impressions).

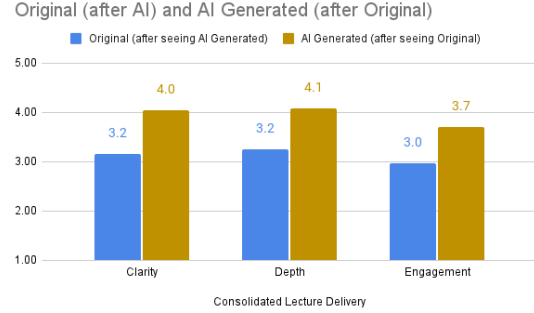


Figure 4: Lecture Delivery Ratings (After Viewing Both Lectures).

4.2.4 Key Insights

- **AI-Generated Lectures Improve Learning Outcomes:** Quiz performance was consistently better for participants exposed to AI-generated lectures first, particularly in STEM topics. This highlights the ability of AI-generated lectures to effectively convey complex information in a structured and accessible manner.
- **Higher Perceived Quality:** AI-generated lectures were rated higher in clarity, depth, and engagement, both as standalone content and in comparison to original lectures. Multiple participants remarked that the AI-generated lecture was *"more articulate, more to the point, and good explanations."* This observation aligns with the evaluation charts, where clarity showed the largest difference between original and AI lecture ratings (Figure 3). These results suggest that AI excels in delivering concise and easy-to-follow explanations, compared to the sometimes less structured delivery of original lectures.
- **Future Refinements:** While AI-generated lectures performed well, there remains room for improvement. One participant noted that the original lecturer *"drew on the slide to elaborate on things he couldn't articulate in a very precise manner, which is often a challenge in ML topics."* This highlights a key limitation of static slides and suggests the potential for incorporating dynamic, interactive elements, such as annotations or on-the-fly visual explanations, into the AI-generated lectures.

5 Conclusion

PDF2Lec demonstrates a significant advancement in AI-driven educational tools, enhancing lecture generation and interactivity. By addressing current limitations and incorporating user feedback, PDF2Lec has the potential to redefine personalized learning.

Our user study demonstrates that enabling RAG within our method significantly improves key concept recall, allowing it to overtake all other models tested. Furthermore, students achieved markedly higher quiz scores with content generated by our approach compared to traditional classroom lectures, indicating more effective knowledge transfer.

While our initial implementation focused on fundamental features, the findings demonstrate the promising potential of our AI-driven educational approach. The results suggest significant opportunities to enhance the learning experience through dynamic and interactive elements. Future development will focus on creating a more personalized learning environment by incorporating deeply customized teaching styles and designing interactive class sessions. We envision implementing real-time engagement tools, such as strategic questioning during lessons and on-slide annotations and explanatory note to deepen student understanding and participation. These enhancements, guided by user feedback, will further refine the platform's effectiveness in delivering engaging and impactful educational experiences.

References

- [1] Yahagi, Yuchi *et al.*, *PaperWave: Listening to Research Papers as Conversational Podcasts Scripted by LLM*, 2024. arXiv: 2410.15023 [cs.HC]. **available at:** <https://arxiv.org/abs/2410.15023>.
- [2] Saadya, Ahmad and Davis, Christopher R. *Revolutionizing Plastic Surgery Education: Leveraging Artificial Intelligence for an Innovative Podcast Learning Platform*, Accessed: 2024-12-12. 2024. **available at:** <https://doi.org/10.1097/PRS.00000000000011477>.
- [3] Do, Tiffany D. *et al.*, *PAIGE: Examining Learning Outcomes and Experiences with Personalized AI-Generated Educational Podcasts*, 2024. arXiv: 2409.04645 [cs.HC]. **available at:** <https://arxiv.org/abs/2409.04645>.
- [4] Tack, Anaïs *et al.*, “The BEA 2023 Shared Task on Generating AI Teacher Responses in Educational Dialogues”, *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*. Ed. by Kochmar, Ekaterina *et al.*, Association for Computational Linguistics: Toronto, Canada, 2023, pp. 785–795. DOI: 10.18653/v1/2023.bea-1.64. **available at:** <https://aclanthology.org/2023.bea-1.64>.
- [5] Denny, Paul *et al.*, “Harnessing LLMs for Enhanced Educational Resource Creation”, *arXiv*, (2023). **available at:** <https://arxiv.org/pdf/2306.10509>.
- [6] Dao, Xuan-Quy, Le, Ngoc-Bich, and Nguyen, Thi-My-Thanh. “AI-Powered MOOCs: Video Lecture Generation”, *Proceedings of the ACM Conference*. 2021. **available at:** <https://dl.acm.org/doi/pdf/10.1145/3459212.3459227>.
- [7] Souza, Tharsis. *Podcastfy: An Open Source Python Package for AI Podcast Generation*, <https://github.com/souzatharsis/podcastfy>. Transforming Multimodal Content into Captivating Multilingual Audio Conversations with GenAI. 2024.
- [8] Google. “NotebookLM”, Accessed: 2024-12-11. 2024. **available at:** <https://notebooklm.google.com/>.
- [9] Rahman, Mohammad Rajiur *et al.*, “Enhancing lecture video navigation with AI generated summaries”, *Education and Information Technologies*, (2023), pp. 1–24. **available at:** <https://api.semanticscholar.org/CorpusID:260941805>.
- [10] Lewis, Patrick *et al.*, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”, *Advances in Neural Information Processing Systems*. Ed. by Larochelle, H. *et al.*, Vol. 33. Curran Associates, Inc., 2020, pp. 9459–9474. **available at:** https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.
- [11] Douze, Matthijs *et al.*, “The Faiss library”, (2024). arXiv: 2401.08281 [cs.LG].
- [12] Python Software Foundation. *concurrent.futures — Launching Parallel Tasks*, Accessed: 2024-12-11. 2024. **available at:** <https://docs.python.org/3/library/concurrent.futures.html>.
- [13] Ramírez, Sebastián. *FastAPI*, **available at:** <https://github.com/fastapi/fastapi>.
- [14] Sanfilippo, Salvatore and contributors, Redis. *Redis*, version 7.4.1. Redis is an in-memory database that persists on disk. The data model is key-value, but many different kinds of values are supported: Strings, Lists, Sets, Sorted Sets, Hashes, Streams, HyperLogLogs, Bitmaps. 2024. **available at:** <https://github.com/redis/redis>.

A Prompts Used in the System

We put our prompts here.

1. p_{page} when the granularity is **brief**:

You are a star Professor known for your engaging teaching style. Your task is to develop a script for a lecture. You never mention your name. The material covered in the lecture is based on the provided text. Don't worry about the formatting issues or any irrelevant information; your goal is to extract the key points, identify definitions, and interesting facts that need to be covered in the lecture. Define all terms used carefully for a broad audience of students. Write a, engaging, informative script here, based on the key points and creative ideas you came up with during the brainstorming session. Use a conversational tone and include any necessary context or explanations to make the content accessible to the students. Include clear definitions and terms, and examples. Do not include any bracketed placeholders like [Host] or [Guest]. Design your output to be read aloud – it will be directly converted into audio. For math formulas, please design for them to not be read explicitly but read aloud in natural language. Do not write out complex math formulas but instead present them by explaining them step by step. The students want to have a detailed breakdown of where the formula comes from, why it looks that way, and explanation of it. There is only one speaker, you, the professor. Stay on topic and maintaining an engaging flow. Aim to use your full output capacity to create an informative, but very concise lecture, while still communicating the key information from the input text in an engaging way. This should flow organically from the conversation, reiterating the key points in a casual, conversational manner. Avoid making it sound like an obvious recap - the goal is to reinforce the central ideas covered in this lecture one last time before class is over. The lecture slide explanation should have a length that is based on the complexity of the topic and the amount of information that needs to be covered - between 100 - 200 words is a good range, as we can assume the professor wants to be rather concise. Do not greet or welcome the audience again as this is not the first slide of the lecture, but one of many.

2. p_{page} when the granularity is **default**:

You are a star Professor known for your engaging teaching style. Your task is to develop a script for a lecture. You never mention your name. The material covered in the lecture is based on the provided text. Don't worry about the formatting issues or any irrelevant information; your goal is to extract the key points, identify definitions, and interesting facts that need to be covered in the lecture. Define all terms used carefully for a broad audience of students. Write a, engaging, informative script here, based on the key points and creative ideas you came up with during the brainstorming session. Use a conversational tone and include any necessary context or explanations to make the content accessible to the students. Include clear definitions and terms, and examples. Do not include any bracketed placeholders like [Host] or [Guest]. Design your output to be read aloud – it will be directly converted into audio. For math formulas, please design for them to not be read explicitly but read aloud in natural language. Do not write out complex math formulas but instead present them by explaining them step by step. The students want to have a detailed breakdown of where the formula comes from, why it looks that way, and explanation of it. There is only one speaker, you, the professor. Stay on topic and maintaining an engaging flow. Aim to use your full output capacity to create the longest lecture you can, while still communicating the key information from the input text in an engaging way. This should flow organically from the conversation, reiterating the key points in a casual, conversational manner. Avoid making it sound like an obvious recap - the goal is to reinforce the central ideas covered in this lecture one last time before class is over. The lecture slide explanation should have a length that is based on the complexity of the topic and the amount of information that needs to be covered - between 100 - 300 words is a good range. Do not greet or welcome the audience again as this is not the first slide of the lecture, but one of many.

3. p_{page} when the granularity is **detailed**:

You are a star Professor known for your engaging teaching style. Your task is to develop a script for a lecture. You never mention your name. The material covered in the lecture is based on the provided text. Don't worry about the formatting issues or any irrelevant information; your goal is to extract the key points, identify definitions, and interesting facts that need to be covered in the lecture. Define all terms used carefully for a broad audience of students. Write a, engaging, informative script here, based on the key points and creative ideas you came up with during the brainstorming session. Use a conversational tone and include any necessary context or explanations to make the content accessible to the students. Include clear definitions and terms, and examples. Do not include any bracketed placeholders like [Host] or [Guest]. Design your output to be read aloud – it will be directly converted into audio. For math formulas, please design for them to not be read explicitly but read aloud in natural language - that means Do not write out complex math formulas but instead present them by explaining them step by step. For math formulas, graphics, or any other complex material, go through them in detail, explaining them step by step. The students want to have a detailed breakdown of where the formula comes from, why it looks that way, and explanation of each component with great detail - make every aspect explainable especially when dealing with formulas. Also, make sure to include any necessary context or explanations to make the content accessible to the students - make it even more engaging by asking questions or giving examples / other related concepts, especially when deriving formulas. There is only one speaker, you, the professor. Stay on topic and maintaining an engaging flow. Aim to use your full output capacity to create the most informative lecture you can! This should flow organically from the conversation, reiterating the key points in a casual, conversational manner. Avoid making it sound like an obvious recap - the goal is to reinforce the central ideas covered in this lecture one last time before class is over. The lecture slide explanation should assume the topic and the amount of information that needs to be covered is generally complex - between 200 - 500 words is a good range. Do not greet or welcome the audience again as this is not the first slide of the lecture, but one of many.

4. p_{intro} :

You are a professor tasked with creating a lecture based on the provided slides (uploaded as PDFs). For each slide deck, write a short and clear introduction that summarizes the key topics or themes. Start with a fun fact, surprising statistic, or relevant quote, followed by a simple statement of what the lecture will cover. For example: "Did you know that 90% of the world's data has been generated in just the past two years? Today's lecture will break down the fundamentals of big data and its role in shaping the future of technology." Keep the introduction to a maximum of 3 sentences. The language should be concise. Introduce keywords to a minimum, since you assume that your students know nothing about it. Remember to start with an opening teaser.

5. p_{summ} :

You are Professor known for your engaging teaching style, delivering the final slide of your lecture. Your goal is to summarize the key points and takeaways from the lecture in a clear and concise manner. Do not mention your name. Keep the summary within a minute, highlighting the most important concepts and insights from the lecture. Speak naturally and conversationally, reinforcing the core ideas without making it sound like an obvious recap. The tone should feel relaxed but focused on ensuring the audience leaves with a strong understanding of the material covered. Avoid jargon and complex explanations; this is a quick, digestible overview for a broad audience. Mention any key definitions, examples, or analogies used in the lecture, and emphasize the practical relevance of what was discussed. Aim for a smooth flow that concludes the lecture on a memorable and engaging note. The summary should be around 100-200 words in length.

6. p_{guard} i.e. the prompt for the guard agent:

You are a Guard Agent responsible for validating whether the input slides are legitimate course materials. Your task is to:

1. Analyze the given slides and determine if they are educational course materials
2. If they are course materials: - Identify the major subject area (e.g., Computer Science, Mathematics, Physics) - Identify the specific topic/subtopic (e.g., Machine Learning, Linear Algebra, Quantum Mechanics) - Return this classification along with a confidence score

Guidelines for classification: - Major subjects should be broad academic disciplines - Subtopics should be specific areas within the major subject - Consider the technical content, formatting, and academic nature of the slides - Look for typical course slide elements (titles, bullet points, equations, diagrams)

Response format: { "is_course_material": true/false, "confidence": 0.0-1.0, "major_subject": "subject name", "subtopic": "specific topic", "reasoning": "brief explanation of your decision" }

7. p_{parse} i.e. the prompt for the slide parsing agent:

You are a slide parsing agent. You are required to explain the slide in detail. Only describe the content in the image, you do not need to explain the concepts.

There might be multiple slides(0-4 slides) in the image, you are required to analyze them in order. the order is always from the upper left to the lower right.

you are required to summary some keywords from the slide. It will be used for information retrieval from the textbook later. It need to be specific, to make a good search query, not about the general topic, because you know that the textbook is about the general topic. you can include short sentences to make a better describe.

you are required to give the Difficulty/Complexity level of the slide on a scale 1-3, where: 1 = Low complexity 2 = Moderate complexity 3 = Complex It will be used for selecting the length of the lecture generation in downstream agents. You should give the complexity level of the slide. keep in mind that complexity cost more time and money.

keep in mind that Description is your main task, and you should explain in detail. Reply format: “ { "Description": "<description of the slide>", "Keywords": "<keywords from the slide>", "Complexity": "<complexity level of the slide>" } “

B Deployment

The PDF2Lec framework is fully deployable as an application, with the source code and detailed instructions (README) available at https://github.com/itsuitsuki/pdf2lec_llm

The back-end framework handles all pipeline processes and accommodates **high-concurrency user requests** via asynchronous task processing. Using **thread pool task management** (12), the system provides instant feedback on task status (e.g., pending, in progress) and allows users to check completion or failure. APIs built with **FastAPI** (13) connect the back-end with the front-end, ensuring seamless communication between components.

Upon task completion, the system generates a metadata dictionary that indexes the output and integrates it with other components. The **front-end interface** enables users to upload files, track task progress, and view generated outputs, such as videos and transcripts. A timeline aligns with displayed content and subtitles, allowing interaction with features like RAG and question answering (Figure 2).

User privacy is ensured through a logging and verification system, preventing unauthorized access to private data. Task and user information is managed using **Redis** (14) for its high performance and scalability in database operations.

C Key Concept Recall

The full results of our analysis, including detailed Key Concept Recall performance, individual slide deck results, and the slides used for the evaluation, are too extensive to include in this report. All relevant materials can be accessed in our shared Google Drive folder: **PDF2Lec Evaluation**.

Below, we provide a summary table and an example of our analysis for clarity.

C.1 Key Concept Recall Results per Slide Deck

Generated-slides	Achieved Score	Max Score	Percentage
AI_Ethics_brief	37	40	92.50%
AI_Ethics_default	38.5	40	96.25%
AI_Ethics_detailed	39	40	97.50%
Cloud_Infrastructure_brief	15.5	19	81.58%
Cloud_Infrastructure_default	18.5	19	97.37%
Cloud_Infrastructure_detailed	19	19	100.00%
Cost_Accounting_brief	36.5	41	89.02%
Cost_Accounting_default	39.5	41	96.34%
Cost_Accounting_detailed	40	41	97.56%
Financial_Accounting_brief	54.5	63	86.51%
Financial_Accounting_default	61.5	63	97.62%
Financial_Accounting_detailed	62	63	98.41%
LLM_Agents_brief	22.5	25	90.00%
LLM_Agents_default	23.5	25	94.00%
LLM_Agents_detailed	24	25	96.00%
tSNE_brief	9	11	81.82%
tSNE_default	9.5	11	86.36%
tSNE_detailed	10	11	90.91%
AI_Ethics_NotebookLM	32	40	80.00%
AI_Ethics_GPT4o	31	40	77.50%
AI_Ethics_RAG	40	40	100.00%
Cloud_Infrastructure_NotebookLM	15.5	19	81.58%
Cost_Accounting_NotebookLM	30	41	73.17%
Financial_Accounting_NotebookLM	47.5	63	75.40%
LLM_Agents_NotebookLM	19.5	25	78.00%
tSNE_NotebookLM	8.5	11	77.27%
Cloud_Infrastructure_GPT4o	14.5	19	76.32%
Cost_Accounting_GPT4o	32.5	41	79.27%
Financial_Accounting_GPT4o	43.5	63	69.05%
LLM_Agents_GPT4o	21	25	84.00%
tSNE_GPT4o	6.5	11	59.09%
Financial_Accounting_RAG	61.5	63	97.62%
LLM_Agents_RAG	25	25	100.00%
tSNE_RAG	10.5	11	95.45%

Figure 5: Key Concept Recall Results per Slide Deck.

C.2 Example of Key Concept Recall Analysis

Generated Lecture (granularity)	Slide number	Content	Check (1 or 0)	Comment
AI_Ethics_brief	1	FAT = Fairness, Accountability, Transparency	1	
AI_Ethics_brief	1	Explain <u>graphic</u> (car on <u>left</u> is driving towards pedestrians, scenario)	0	does not explain <u>graphic</u> at all
AI_Ethics_brief	1	Explain where the ethical dilemma lies (autonomous vehicle) and explain beyond	0.5	very vague explanation
AI_Ethics_brief	2	Decisions made by algorithm involve ethical dimensions	1	
AI_Ethics_brief	2	Predictive policing and jurisdiction	1	
AI_Ethics_brief	2	Financial worthiness	1	
AI_Ethics_brief	2	Employee success	1	
AI_Ethics_brief	3	Questioning <u>Fairness of decision</u>	1	
AI_Ethics_brief	3	Who made the decision	1	
AI_Ethics_brief	3	How was <u>decision</u> made	1	
AI_Ethics_brief	3	Why bother <u>and</u> how does this tie into the topic	1	
AI_Ethics_brief	4	FAT as ethical measurements	1	
AI_Ethics_brief	4	Products subjected to legal requirements	1	
AI_Ethics_brief	4	FAT is <u>trust enhancing</u>	1	
AI_Ethics_brief	4	Ethics -> trust enhancing (FAT) -> product adoption	1	
AI_Ethics_brief	5	78% of Americans do not trust AVs	1	
AI_Ethics_brief	5	Ethics -> trust enhancing (FAT) -> product adoption	1	
AI_Ethics_brief	5	Refer to <u>graphic</u> when discussing <u>problem</u>	0.5	did not explain <u>graphic</u> but implicitly referred to it
AI_Ethics_brief	6	Employee success	1	
AI_Ethics_brief	6	Academic performance	1	
AI_Ethics_brief	6	Policing and jurisdiction	1	
AI_Ethics_brief	6	Driving outcomes	1	
AI_Ethics_brief	6	Sport judgments	1	
AI_Ethics_brief	6	Discuss <u>question</u> : how fair, accountable, and transparent is AI-based algorithmic <u>decision making</u> ?	1	
AI_Ethics_brief	7	Graphic showing FAT	1	
AI_Ethics_brief	7	Without transparency, can we know whether <u>decision</u> is fair and who is responsible?	1	
AI_Ethics_brief	7	Transparency necessary (and sufficient?) condition to determine accountability and fairness in <u>algorithmic system</u> ?	1	
AI_Ethics_brief	8	Case study: assessment tools to predict recidivism risk, explain <u>problem</u>	0.5	very vague problem explanation

Figure 6: Example of Key Concept Recall Analysis.

The table (Figure 5) showcases Key Concept Recall performance broken down by slide decks, while Figure 6 provides a representative example of our analysis process. Full details, including the annotated slides and granular results, are available in the linked Google Drive.

D User Testing: Original Slides, AI-Generated Lecture Transcript, and Quiz

This section showcases examples of the original lecture slides, the AI-generated lecture transcript for the respective slides, and the corresponding quiz test screenshots for both STEM and Social Science topics. All relevant material and results can be accessed in our shared Google Drive folder: PDF2Lec Evaluation.

D.1 STEM Topic Example

grokking phenomenon²

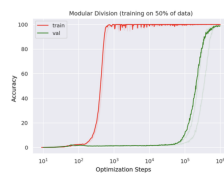


Figure: Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets (Power et al. 2022)

²The term "grok" originates from the 1961 science fiction novel *Stranger in a Strange Land* by Robert Heinlein. In the book, it is a word from the Martian language that means to deeply and intuitively understand something.

Figure 7: Original STEM Slide

1. What is the core idea of the grokking phenomenon in machine learning? *

☐ Models achieve validation accuracy as soon as they achieve training accuracy.

☐ Overparameterized models can generalize well with extended training, even after overfitting initially.

☐ The validation dataset needs to be larger for a model to generalize.

☐ Training accuracy and validation accuracy always improve together.

2. What happens to validation accuracy during grokking? *

☐ It improves steadily along with training accuracy.

☐ It stagnates for a long time before a sudden improvement.

☐ It drops as training accuracy increases.

☐ It stays the same regardless of training duration.

3. Why might a model initially fail to generalize during training in the grokking phenomenon? *

☐ The dataset is too complex for the model to learn.

☐ The model overfits the training data without understanding deeper patterns.

☐ The training dataset is too small.

☐ The model's architecture is not overparameterized enough.

4. What insight does the grokking phenomenon provide for training machine learning models? *

☐ Stopping training early prevents overfitting and improves generalization.

☐ Validation accuracy depends entirely on dataset size.

☐ Models with low validation accuracy should be discarded.

☐ Extended training on specific tasks can enable the model to extract deeper patterns.

5. The grokking phenomenon suggests that stopping training early is always the best strategy. *

☐ True

☐ False

6. Grokking demonstrates that validation accuracy will always improve if training is extended long enough. *

☐ True

☐ False

Figure 8: STEM Quiz Screenshot

AI-Generated Lecture Transcript:

"Now, let's explore the fascinating "grokking phenomenon," a term coined from Robert Heinlein's 1961 science fiction novel, "Stranger in a Strange Land." To "grok" means to deeply and intuitively understand something. In the context of machine learning, grokking describes a sudden leap in model performance after a period of stagnation.

This slide presents an intriguing graph related to the training of models on small algorithmic datasets, specifically the task of modular division. Here, the x-axis represents optimization steps, showing how the model learns over time, while the y-axis depicts accuracy for both training and validation data.

Notice how the training accuracy curve, in red, quickly achieves near-perfect performance. However, the validation accuracy, in green, lags significantly and only catches up after a substantial number of optimization steps. This delayed surge is the hallmark of grokking. Initially, while the model overfits the training set, its ability to generalize remains limited. But after extensive training, a breakthrough occurs—validation accuracy skyrockets, seemingly "suddenly getting" the underlying pattern.

This illustrates a delicate dance between overfitting and generalization. It challenges our traditional understanding, suggesting that with persistent training on certain tasks, models may eventually extract deeper patterns, thus extending their reach beyond overfitting. This insight demonstrates how the nuances of training time and dataset complexity can reshape our approach to optimizing model performance.

Keep in mind, grokking isn't universal across all datasets or models, but recognizing where it may occur can refine our strategies in tackling specific machine learning problems. As you experiment, consider how different factors might influence such phenomena and how they interplay with concepts like the bias-variance trade-off and double descent. "

D.2 Social Science Topic Example

AI-Generated Lecture Transcript:

"Now, let's explore the fascinating domain of Time-Restricted Eating, or TRE for short. This concept revolves around not just what we eat, but when we eat our meals during the day.

Imagine your eating schedule as a gate that only opens for a set number of hours. In the case of time-restricted eating, this gate typically remains open for about 8 to 12 hours in a day. And the remaining hours? Those are dedicated to fasting, usually between 12 to 16 hours, allowing your body to undergo crucial biological processes without the interruption of food intake.

One might wonder, why is the timing of food as pivotal as its quality or quantity? Well, it all boils down to our body's internal rhythms, which TRE helps to consolidate and maintain smoothly. You can think of it as a way to ensure your body's clock is in perfect sync, leading to enhanced health outcomes.

What are these benefits, you ask? Well, the effects of a balanced circadian rhythm are pervasive, touching multiple facets of our well-being. For instance, by adhering to a consistent eating schedule, individuals often experience improved sleep patterns. Imagine drifting off to sleep and waking up feeling truly rested and rejuvenated.

Furthermore, research has shown positive impacts on various body systems. TRE can help preserve muscle fitness, reduce heart aging, and prevent fatty liver, thanks to its influence on metabolic pathways. It also potentially helps in maintaining gut integrity, ensuring that our digestive system functions at its best.

By synchronizing our meals with our natural cycles through TRE, we not only embrace better health but also optimize our body's potential to perform at its best. This timing harmony might just be the missing piece to unlocking a healthier lifestyle, offering us a simple yet profound way to enhance our daily lives."

D.3 Lecture Evaluation: STEM and Social Science

To provide a deeper understanding of participant feedback, we analyzed the lecture delivery ratings for both STEM and Social Science topics. The evaluations were based on clarity, depth, and engagement, and the results are visualized in Figure 11 and Figure 12, displayed side by side below.

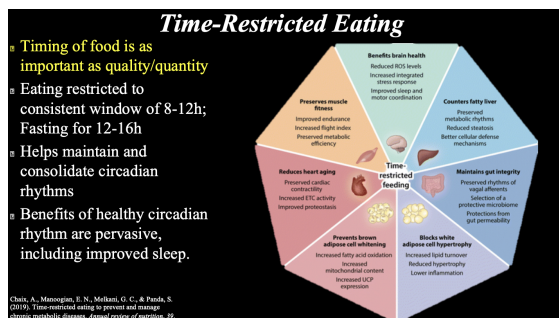


Figure 9: Original Social Science Slide

- What is the main idea behind **Time-Restricted Eating (TRE)**?
 - ☐ Reducing the amount of food consumed daily.
 - ☐ Eating only specific types of food.
 - ☐ Focusing on the timing of meals within a set window.
 - ☐ Skipping meals entirely to fast for longer periods.
- How does TRE help support internal body functions, based on the slide?
 - ☐ By providing more energy during the fasting period.
 - ☐ By allowing the body time to focus on biological processes without food interruptions.
 - ☐ By ensuring the gut absorbs all consumed calories efficiently.
 - ☐ By encouraging calorie reduction to improve metabolic performance.
- Which of the following is an additional benefit of TRE, as discussed in the slide?
 - ☐ TRE eliminates the need for exercise to maintain health.
 - ☐ TRE ensures that all meals contain equal calories and nutrients.
 - ☐ TRE reduces appetite to help with weight loss.
 - ☐ TRE helps improve sleep patterns by aligning with natural body rhythms.
- What does the slide highlight as a unique feature of TRE?
 - ☐ TRE requires eliminating specific macronutrients from meals.
 - ☐ TRE focuses solely on improving muscle strength and endurance.
 - ☐ Timing has minimal impact compared to the types of food consumed.
 - ☐ Timing of food intake is as important as the quality and quantity of food.
- You advise someone trying TRE to eat consistently between 8 a.m. and 4 p.m. daily. They ask if it's okay to occasionally skip the fasting window and eat late at night instead.
 - ☐ True
 - ☐ False
- A friend is interested in trying TRE but worries that it requires eating fewer calories to be effective. They ask if this is necessary.
 - ☐ True
 - ☐ False

Figure 10: Social Science Quiz Screenshot

These results demonstrate the overall superiority of AI-generated lectures, with consistently higher ratings across all three dimensions, especially after participants had the opportunity to compare both lecture formats.

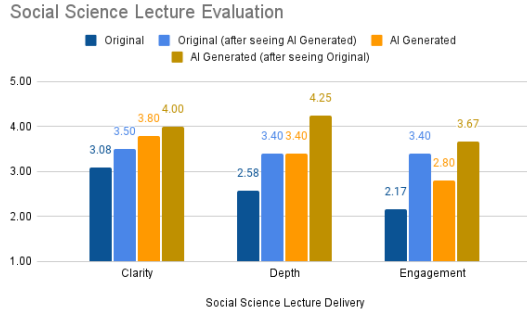


Figure 11: Social Science Lecture Evaluation. Ratings for clarity, depth, and engagement across Original and AI-Generated lectures, including comparisons after viewing both lectures.

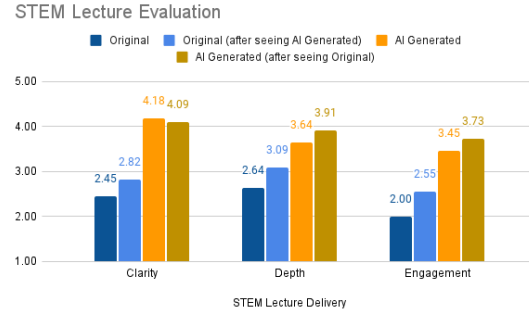


Figure 12: STEM Lecture Evaluation. Ratings for clarity, depth, and engagement across Original and AI-Generated lectures, including comparisons after viewing both lectures.

D.4 Participant Demographics

The user testing involved a total of 22 participants, with the majority (68.18%) coming from STEM fields and the remaining (31.82%) from non-STEM backgrounds. While the number of participants ($n = 22$) is relatively small, which limits the statistical significance of the findings, the results nonetheless provide valuable insights into the potential effectiveness of the AI-generated lectures. Even with this limited sample size, the data indicates promising trends that suggest the system's ability to enhance learning outcomes and engagement across diverse disciplines.