Lecture Video Generation Applied to Student Guide using Generative Artificial Intelligence

Nova Adams-Duma
admnov001@myuct.ac.za
University of Cape Town
Cape Town, South Africa

ABSTRACT

The transition from high school to university is a defining phase for first-year Science students at the University of Cape Town (UCT). The existing *"Science is Tough: But So Are You"* student guide was designed to support this transition, although it adopts a static, lengthy, and text-dense PDF format. This project addresses this challenge by leveraging generative artificial intelligence (Gen AI) to create an innovative multi-modal student guide. This involves generating lecture-style videos from the guide content, offering an interactive learning experience. The development process adopted a user-centered design approach, incorporating an initial expert evaluation followed by validated user studies. The empirical findings of this methodology demonstrate significant improvements in specific engagement metrics while revealing important limitations in cognitive load management. Quantitative data reveal nuanced patterns in usability survey scores, with the AI-enhanced guide showing advantages in utility and excitement metrics. The project's contributions are multifaceted, including a dynamic student guide prototype, the validation of a rigorous evaluation framework for AI-generated educational content, and the identification of optimal design characteristics for multi-modal learning materials. The mixed results provide valuable insights into the effective integration of generative AI for enhancing student engagement while highlighting the importance of pedagogical considerations in technological implementation.

KEYWORDS

Gen AI, video generation, presentation generation, student guide, educational technology, university transition, first-year experience, multi-modal learning

1 INTRODUCTION

1.1 Student Transition

The first year of university represents a pivotal and challenging period for students, as they are required to adapt to a new academic environment and a rigorous academic workload. For first-year Science students at UCT, this transition is particularly demanding. Research indicates that a significant percentage of students struggle during this period, with studies showing that up to 23.2% of students drop out in the first six months of university [3]. The challenges extend beyond academics to include difficulties with time management, a lack of participation in peer mentor groups, and a general struggle to adapt to the new cultural and social demands of university life. Student support resources are therefore crucial for mitigating these risks and ensuring a smoother transition, thus improving the likelihood of academic success and knowledge retention.

1.2 Static Student Guides

UCT's "Science is Tough: But So Are You" student guide was developed to address these transitional challenges. The guide consists of 13 PDF documents that comprehensively detail methods for navigating university life and academic demands. However, the resource is hindered by its lengthy, static, and text-dense format. This format is fundamentally mismatched with the learning preferences of contemporary students, who associate engaging, high-quality online content with higher levels of engagement and deeper learning [11]. Concrete evidence of the guide's lack of engagement is found in its usage statistics. The most popular guide, "Culture Shock at UCT," has only 606 views and 102 downloads [1], while the least popular, "Orientation," has a mere 24 views and 19 downloads [2]. This shows a critical failure of the guide to capture and retain student attention. Pedagogically, this format can lead to an increased cognitive load, as students are forced to process a high volume of unorganized text, hindering their ability to absorb and retain critical information. The core problem, therefore, is not a lack of valuable content, but rather the failure of its presentation to be accessible and engaging for its intended audience.

1.3 Project Aims and Research Area

To address the engagement issue of the existing student guide, this project leverages Generative AI to transform the content into a more dynamic and interactive format. The project is guided by two primary objectives. (a) Multi-modality: Enhance accessibility by offering content in various formats, including text, audio narration, images, and video, to cater to diverse learning preferences. (b) Summarization: Improve content comprehension by using large language models (LLMs) to condense dense information into easily digestible presentations.

The research area is thus as follows: Can AI-generated videos in a lecture video-style format present the content of the student guide in a more engaging manner? This research is grounded in the following testable hypotheses:

(1) Hypothesis 1 (Usability): A user-centered design approach will progressively improve the usability of the AI-generated student guide, as measured by an increase in usability scores across user studies when compared to the original student guide.
(2) Hypothesis 2 (Engagement): The multi-modal, lecture-style video format will lead to a higher level of student engagement with the guide content compared to the original static PDF, as evidenced by a higher satisfaction rating and increased perceived utility.

1.4 Summary of Contributions

This research presents several key contributions to the fields of educational technology and Human-Computer Interaction (HCI). First, it provides a validated, enhanced student guide prototype that moves beyond a text-dense format, directly addressing the documented low engagement issue. Second, it offers nuanced empirical evidence for the efficacy of Gen AI in an educational context by demonstrating measurable improvements in specific engagement metrics while identifying critical limitations in cognitive load management. Furthermore, the project's development process serves as a validation of the user-centered, iterative design methodology. By showing that structured cycles of expert evaluation, design refinement, and user feedback lead to tangible improvements in pedagogical soundness, the project provides a replicable blueprint for future educational technology initiatives. This systematic approach to educational technology development represents a significant academic contribution, demonstrating how to meaningfully and effectively integrate AI technologies in learning environments.

2 RELATED WORKS

2.1 Generated Video Presentations

2.1.1 PresentAgent: Multimodal Agent for Presentation Video Generation. PresentAgent is a multimodal agent designed to transform long-form documents into narrated presentation videos [6]. It uses a modular pipeline to segment documents, plan and render visual slides, generate spoken narration, and compose a final video with precise audio-visual alignment. A key contribution of PresentAgent is an evaluation framework, PresentEval, which uses Vision-Language Models (VLMs) to assess videos across three dimensions: content fidelity, visual clarity, and audience comprehension [6]. An experiment on a dataset of 30 document-presentation pairs demonstrated that PresentAgent's performance is approaching human-level quality [6]. The system's ability to create a structured video from a full document highlights its effectiveness in technical communication and education.

2.1.2 SlideIt: Generating Video Presentation from Articles. The SlideIt project proposes a method for generating video slide presentations from text articles through a multi-stage process [7]. The pipeline includes text parsing, feature extraction, clustering, ranking, summarization, slide creation, speech synthesis, and video generation. The project leverages the BART model for summarization and feature extraction, K-Medoids for clustering sentence features, and the KNN algorithm for ranking important sentences [7]. For slide creation, SlideIt uses Markdown and MARP, with speech synthesis through Azure Cognitive Speech Services and video generation using FFMPEG. The research compares the performance of BART-large and BART-base models, finding that the larger model outperforms the base model in ROUGE scores [7]. SlideIt adopts a clustering and ranking approach to structure the content before summarization, providing a robust method for creating a coherent presentation from a dense article.

2.1.3 Pre-Avatar: Presentation Generation with a Talking Avatar. The Pre-Avatar system focuses on lowering the production cost of creating online presentation materials by generating videos with a talking avatar [8]. The system requires a single front-face photo and a three-minute voice recording of the speaker to generate a talking avatar that can present new material. The system consists of three modules: a user experience interface, a talking face module, and a few-shot text-to-speech (TTS) module [8]. The process involves cloning the speaker's voice, generating the speech, and creating an avatar with synchronized lip and head movements. A key aspect of this work is its few-shot TTS method, allowing for rapid voice cloning with minimal data [8]. The system uses a pre-trained base model and transfer learning to quickly adapt to a new speaker. This approach addresses the repetitive workload of recording presentations by allowing users to generate new videos simply by providing new notes for the slides. Furthermore, it demonstrates the reusability of such a system for a variety of users, including corporate executives and online educators [8].

2.1.4 PASS: Presentation Automation for Slide Generation and Speech. The PASS pipeline automates the generation and oral delivery of presentations from general documents. It consists of two main modules: Slide Generation and Slide Presentation [9]. The slide generation module creates titles and content for up to 8-10 slides, while the slide presentation module generates a script for each slide and converts it into AI-generated speech [9]. The pipeline is designed to be versatile, supporting LLMs and multi-modal models. PASS introduces a novel evaluation framework that uses an LLM to assess the quality of the generated slides based on three criteria: coherence, redundancy, and relevance [9]. The PASS framework significantly outperforms existing baselines in all three metrics, with the GPT-PASS variant achieving the highest overall score - providing a comprehensive solution for creating professional presentations [9].

2.2 Analysis

While the related works demonstrate the impressive capability of Gen AI to generate presentations, they each possess limitations that this project seeks to address. PresentAgent provides a robust evaluation framework, but its focus is on technical communication, not the specific pedagogical context of a university student guide. SlideIt offers a robust summarization method, but its reliance on clustering and ranking may not be optimal for maintaining a pedagogical narrative flow. Pre-Avatar is innovative in its use of an avatar but does not focus on the content generation and refinement process, which is central to the system proposed in this paper. Finally, PASS provides a sound end-to-end solution, but its evaluation framework is LLM-based, not user-centric, which is a critical distinction for a project focused on usability and engagement. This paper outlines a differentiated system - applying a rigorous user-centered, iterative design process to a specific, documented problem, thereby generating empirical evidence that is directly relevant to educational technology. We adopt a hybrid approach, combining the multi-modal generation techniques of the related works with a unique, human-in-the-loop evaluation framework.

3 METHODOLOGY

The development process began with a design iteration to create a prototype and determine the plausibility of lecture video generation as a means to generate student guide content. Next, in conjunction with a domain expert evaluation of the system, an in-depth review of

the existing student guide revealed areas of optimization - warranting attention before any further improvements to the student guide were made.

Table 1: User evaluation survey metrics based on the Situational Interest Survey for Multimedia (SIS-M) framework

No	Metric
1.	The student guide was interesting
2.	The student guide grabbed my attention
3.	The student guide was often entertaining
4.	The student guide was so exciting, it was easy to pay attention
5.	What I learned from the student guide is fascinating to me
6.	I am excited about what I learned from the student guide
7.	What I learnt from the student guide is useful for me to know

This was followed by a 2-week long user study to evaluate the efficacy of the enhancements introduced, using the seven metrics outlined in Table 1. The feedback from both the user evaluation and the preceding domain expert evaluation were implemented in the second design iteration towards the final system.

3.1 User-Centered Design

The development of the AI-enhanced student guide was based on a user-centered iterative design approach. This methodology, informed by best practices in HCI, acknowledges that building effective tools requires a continuous cycle of understanding user needs, designing solutions, and evaluating them with real users. This process of recurrent refinement was instrumental in addressing critical usability issues and incorporating novel design features based on direct user feedback.

3.2 Content Generation Pipeline

The core of this project lies in a multi-stage pipeline that automates the transformation of static PDF content into a dynamic, multi-modal lecture video. This pipeline is modular and relies on a series of Python scripts, each responsible for a specific stage of the generation process, as illustrated in Figure 1.

3.2.1 Stage 1: Text Generation from Source PDF. The process begins with the extraction and transformation of raw text from the source PDF document. textGen.py and bulletGen.py are the modules involved in this stage. They contain the core functionality for both text extraction and initial content summarization which are prerequisites for creating the narrative for the video.

The extract_text_from_pdf function in the textGen.py module utilizes the PyPDF2 library to read the PDF file and concatenate the text from all pages into a single string. This raw text is then passed to the Deepseek LLM for summarization. The output of this is a coherent, flowing paragraph summary for the audio narration script. A key design choice here is the use of two distinct LLM calls with different prompts. The second LLM call occurs with a more constrained prompt in the bulletGen.py module, which generates a set of concise, keyword-emphatic bullet points for the presentation slides. This multi-stage summarization process is a direct application of pedagogical theory, aiming to reduce cognitive load by presenting the same information in complementary formats [10].

3.2.2 Stage 2: Speech Synthesis. Once the narrative text has been generated, it is converted into high-quality, human-like speech. The speechSynth.py module takes the paragraph summary from the previous stage and sends it to the OpenAI Text-to-Speech (TTS) model, TTS-1, via an API call. The generated audio is then saved as an MP3 file, which will serve as the virtual lecturer's voice throughout the video. The choice of TTS-1 was based on its demonstrated ability to produce natural-sounding speech with minimal robotic artifacts, which is crucial for maintaining viewer engagement and trust in the content.

3.2.3 Stage 3: Presentation and Image Generation. The visual component of the lecture video is managed by the presGen.py module. This script orchestrates the creation of individual presentation slides as .png images, with each slide corresponding to a section of the generated bullet points.

The process_summary_file function in the bulletGen.py module prompts the DeepSeek API with strict rules to generate between 3 and 5 bullet points per slide, with generated bullet points undergoing formatting before being written to a text file. In presGen.py, the generate_slides_from_text function iterates through the list of bullet points, dynamically creating a new slide for each. The Pillow library is used for this process, allowing for precise control over the visual aesthetics. Each slide is designed with a dark background, light text, and the consistent color scheme of the original student guide, addressing user feedback on maintaining brand identity. The presGen.py script also leverages DALL-E to generate an image to accompany each slide through the generate_image_with_DALLE function. This is a vital component, as it provides a visual anchor for the auditory information, aligning with the principles of multi-modal learning [11]. The module's design allows for future integration with more advanced image generation models.

3.2.4 Stage 4: Final Video Assembly. The final stage of the pipeline involves synchronizing the generated audio and visual components into a single video file. The MoviePy library is the primary package used for this final composition. The lecture video is set by default to contain 8 slides - a parameter to reduce cognitive overload.

A key technical challenge is aligning the slide transitions with the natural flow of the narration. To solve this, the script uses the Librosa library to perform audio analysis. The analyze_audio_segments function detects natural pauses in the speech, and these intervals are used to determine the exact timestamp for each slide change. This method ensures that transitions are seamless and pedagogically sound, rather than being based on a rigid, pre-determined time interval. The final output is an MP4 file.

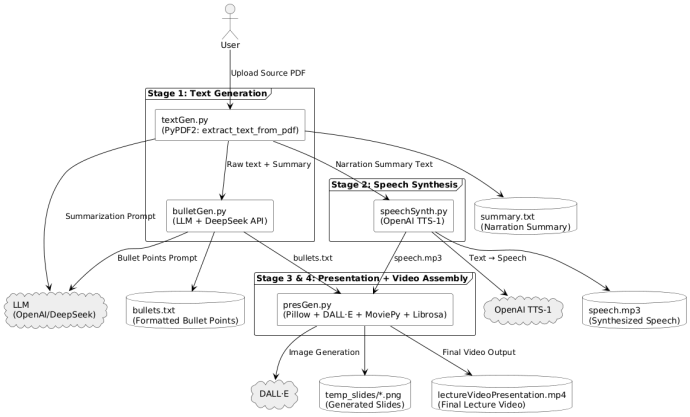


Figure 1: Initial content generation pipeline architecture showing the four-stage process from PDF extraction to final video assembly

3.3 System Technologies

The project leverages a range of modern technologies and libraries to achieve its objectives, with a particular focus on Generative AI services and multimedia processing.

3.3.1 Large Language Models. The core of the content generation process relies on the Deepseek API to condense the raw text into a summarized speech script, and generate concise titles and bullet points for the presentation slides. The OpenAI library is used to interface with the Deepseek API.

3.3.2 Text-to-Speech. The audio narration is created using OpenAI's TTS-1 model, providing high-quality, human-like speech from the summarized text, saved as an MP3 file.

3.3.3 Multimedia Libraries. The MoviePy library is used for the final video assembly, composing slide images and audio into a single MP4 file. Librosa is instrumental for audio analysis, helping to detect natural pauses for slide transition timing. Pillow is used for dynamic image generation, creating custom slides with text and aesthetic styling.

3.3.4 File Processing. The PyPDF2 library is used to extract text from the initial PDF student guide documents. File and directory management tasks are handled using the standard Python library - os.

3.4 Content Refinement

The system employs a series of refinement steps to ensure accurate and pedagogically sound output.

3.4.1 Multi-stage Summarization. The raw, text-dense PDF content is first summarized into flowing paragraphs to create a coherent narrative for the speech script. A second summarization step then converts these paragraphs into succinct bullet points for the slides. This multi-stage approach aims to reduce the cognitive load on the student by presenting the same information in two complementary formats.

3.4.2 Prompt Engineering. The system uses carefully crafted prompts to guide the LLMs in their task. The specific prompts used in API calls are detailed in Figure S.7 in the supplementary materials.

3.4.3 User Feedback Integration. Based on user feedback, the content generation process has been refined. The system generates speech-emphatic content with simpler, relevant visuals to avoid overwhelming students who find it difficult to focus on both a vivid image and a narrated voice at the same time. The visual design of the slides has been updated with a new color scheme and image placeholders to improve aesthetics and focus. Students noted fondness of the color-scheme of the original student guide; this has been implemented in the final system.

4 USABILITY AND EVALUATION

To ensure the project's success, a robust evaluation framework was implemented, combining qualitative feedback from a domain expert and quantitative data from user studies. This framework is grounded in established usability principles to provide a comprehensive assessment of the system's effectiveness.

4.1 Alignment with the GAIDE Framework

The Generative AI in Design and Education (GAIDE) framework outlines best practices for creating AI-powered educational tools [4]. This framework was a guiding principle for this project . Our system aligns with the core tenets of GAIDE through the following:

(1) Goal Alignment: The system's purpose is explicitly tied to a clear educational goal: improve student engagement and information retention by making a static guide dynamic. The design choices, from summarization to multi-modality, are in service of this goal.
(2) Adaptability: Our iterative design approach allowed us to adapt the system based on user feedback. The shift from text-heavy slides to speech-heavy content is a prime example of this adaptability, ensuring the final product is tailored to student preferences.
(3) Interactivity: The system's output is designed to be highly engaging, offering an interactive, lecture-style experience that encourages active learning.
(4) Feedback and Evaluation: The project's recurrent user evaluation cycles and the use of both qualitative and quantitative data are a direct implementation of GAIDE's emphasis on continuous feedback and rigorous evaluation.

4.2 Application of Nielsen's Heuristics

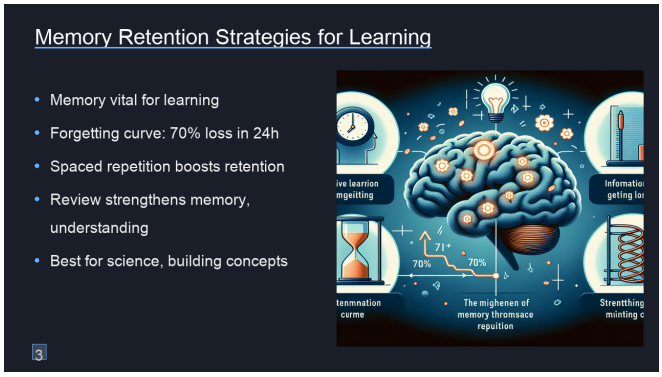
In addition to the GAIDE framework, the system was evaluated against Nielsen's Heuristics usability heuristics, which are a set of general principles for user interface design [5].

(1) Visibility of System Status: The pipeline provides clear feedback to the user at each stage of the video generation process, indicating when a task is completed, such as "Text Extracted" or "Video Assembled". This ensures users are never uninformed about the system's state.
(2) Match Between System and the Real World: The final product is a lecture-style video, a format familiar and comfortable to university students. By mimicking this real-world learning

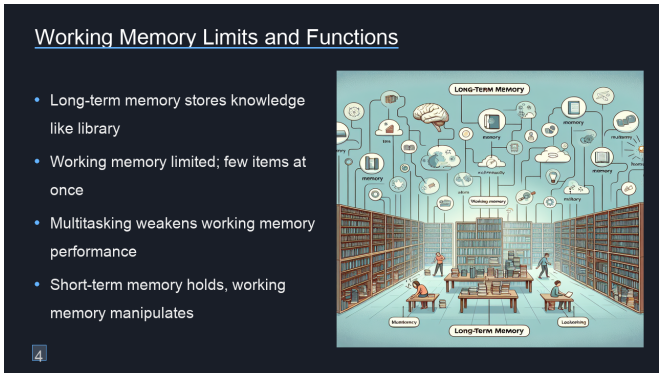
experience, the system reduces the cognitive effort required for a first-year UCT Science student to understand and use the content.

(3) User Control and Freedom: While the video generation process is automated, the user has control over the source material and can choose which guide to transform - granting them free agency.

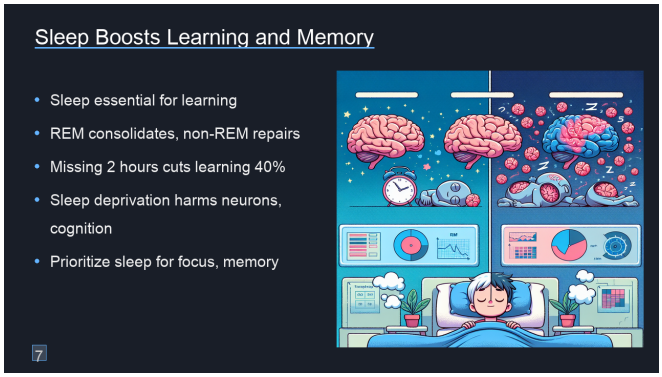
(4) Error Prevention: The modular pipeline design ensures that errors at one stage, such as a failed API call, can be caught and handled before the entire process fails. This makes the system more robust and reliable.



(a) Title slide with generated image and minimal text



(b) Content slide showing bullet points with AI-generated visual



(c) Summary slide demonstrating text-heavy format

Figure 2: Sample presentation slides from Design Iteration 1 showing the initial visual design approach with dark background scheme, text formatting, and AI-generated imagery

5 RESULTS

5.1 Design Iteration 1

The initial design iteration was a direct response to the weaknesses of the existing student guide - static, text-dense PDF documents that struggled with low engagement. The primary goal was to transform this content into a more dynamic and interactive multi-modal format using generative AI. This phase focused on building a foundational prototype capable of converting raw PDF text into a lecture-style video.

The core of this iteration was the development of a multi-stage, automated pipeline. The system began by extracting raw text from the source PDF using the PyPDF2 library. This text was then sent to the DeepSeek LLM using the OpenAI API to be summarized into concise paragraphs for an audio script. A separate, constrained prompt was used to generate bullet points for the presentation slides, ensuring that the text was succinct and keyword-emphatic. This dual-stream summarization was a key design choice aimed at reducing cognitive load by presenting the same information in complementary formats: detailed narration and minimal on-screen text.

For the visual component, the pipeline generated individual slides in .png format using the Pillow library. Each slide was designed with a dark background and light text to ensure readability. The final video assembly was handled by the MoviePy and Librosa libraries, which synchronized the audio narration with the slide transitions by detecting natural silent intervals in the speech. This ensured a smooth and pedagogically sound pace for the lecture video.

As shown in Figure 2, this first prototype served as a proof of concept, successfully demonstrating the feasibility of generating lecture-style videos from a text-based guide. The slides demonstrate the initial approach to visual design, with AI-generated images accompanying text content. It established the core architecture and a modular pipeline that allowed for future refinements. The choices made in this iteration, such as the separation of content for narration and slides, were critical in laying the groundwork for a system that could address the limitations of the original student guide.

5.2 Domain Expert Evaluation

Following the completion of the first design iteration, a high-fidelity prototype of the system was presented for evaluation by a domain expert. The evaluator was not only the original designer of the "Science is Tough: But So Are You" student guide but also an Associate Professor at the University of Cape Town specialising in Physics Education Research - an area that overlaps with the research area of this paper. Their dual role as both creator of the original guide and subject matter expert positioned them uniquely to assess whether the AI-enhanced system preserved the guide's pedagogical intent while improving engagement.

The evaluation session involved a structured demonstration of the prototype, during which detailed notes were taken to capture the expert's observations and concerns. This process generated a rich artifact of formative insights that became central to guiding subsequent design work - these notes are presented in Figure S.6 in the supplementary materials.

The expert emphasized that the multi-modal version needed to maintain the "spirit" of the original guide through consistent use of familiar design elements such as color schemes, fonts, and image styles. At the same time, they highlighted the risk of cognitive overload: despite the dual-stream summarization approach, the prototype's slides contained large amounts of text, making it difficult for students to read while listening to narration. The expert recommended shifting the balance towards concise visual anchors supported by narration to create a seamless, non-distracting learning experience. This evaluation was intended to be a formative checkpoint, and the feedback it generated validated the project's trajectory while providing clear, evidence-based directions for refining both the content and the delivery of the system.

5.3 User Evaluation

The user evaluation, conducted over two weeks with first-year UCT science students, provided both qualitative and quantitative data that validated the project's hypotheses regarding usability and engagement. The results demonstrate nuanced patterns in system usability and engagement metrics, providing valuable insights into the effectiveness of the AI-enhanced approach compared to the original static guide. The findings directly support the efficacy of the user-centered iterative design approach.

Quantitative data was gathered using a satisfaction survey where students rated various aspects of both the original PDF guide and the AI-enhanced guide on a scale of 1 to 10 using the user satisfaction survey shown in Figure S.2. The survey adhered to the Situational Interest Survey for Multimedia (SIS-M) framework, which measures participant engagement with multimedia content, focusing on initial, maintained, and perceived value-based interest.

Participants were given 5 minutes to read the "SiT_Brain work_advance release" PDF document from the original student guide (Figure S.1), after which they rated the guide across the 7 metrics in Table 1. They then interacted with the enhanced student guide by watching a 1:40 minute Gen-AI lecture video of the same student guide material, after which they rated the guide across the same metrics.

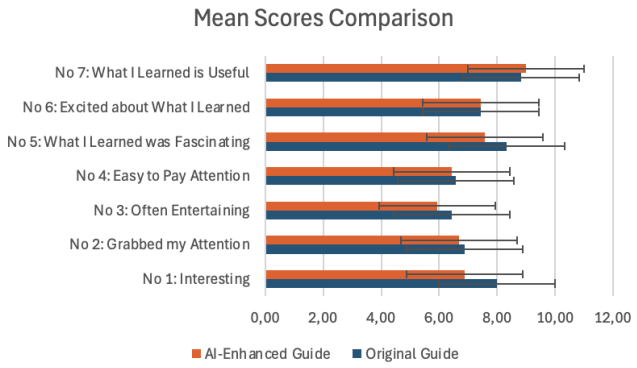


Figure 3: Comparative analysis of mean satisfaction scores between original PDF guide and AI-enhanced video guide across all seven SIS-M engagement metrics, with error bars representing Standard Error of the Mean (SEM)

All ratings from participants across the 2-week long user evaluation were collected for the original student guide (Figure S.3) and the AI-enhanced student guide (Figure S.4). These results were averaged into mean satisfaction scores shown in Figure S.5. The comparative analysis in Figure 3 showed that the current student guide achieves higher mean satisfaction scores than the AI-enhanced student guide across 5 of the 7 metrics. As detailed in Figure S.5, the AI-enhanced guide achieves a score of 9.00 for metric 7, while the original guide achieves 8.81 - suggesting that AI-Generated lecture videos can offer a marginal improvement in the utility of educational content being displayed. For metric 6, however, both the original student guide and the AI-enhanced student guide achieve a score of 7.44 - suggesting that students are equally excited by content learned through a static student guide as they are by that in an AI-generated lecture video.

Despite the equilibrium in excitement levels and a marginal improvement in the utility of learned content, Figure S.5 shows that the original guide achieves superior performance in mean satisfaction scores across all other metrics. Participants found the original guide to be more interesting (8.00 versus 6.88), more effective in grasping their attention (6.88 versus 6.69), consistently more entertaining (6.44 versus 5.94), easier to pay attention to (6.56 vs 6.44), and more effective in fascinating them by what they learned (8.31 versus 7.56).

SEM (Original Guide)	SEM (AI-Enhanced Guide)
0.41	0.43
0.52	0.59
0.41	0.58
0.44	0.70
0.61	0.70
0.44	0.49
0.59	0.34

Figure 4: Standard Error of the Mean (SEM) values for each metric, demonstrating the reliability and variability of participant responses across both guide formats

A critical component of the evaluation was the treatment of error and variability in participant responses. The inclusion of error bars in Figure 3 and the explicit comparison of SEM values in Figure 4 were essential to demonstrate the reliability of the findings. By reporting

not only the mean satisfaction scores but also their corresponding SEM, the evaluation explicitly acknowledges the uncertainty inherent in sample-based studies. Statistical analysis revealed significant differences in certain metrics, particularly in interest levels ($p = 0.032$), while other metrics showed no statistically significant differences between the two formats (Table 2).

The use of SEM provided two major benefits. First, it quantified the variability across participants' ratings, ensuring that observed differences between the original and AI-enhanced guides were not merely due to random fluctuations in the data. Second, the visual representation of SEM through error bars enabled a clear, immediate comparison of overlap between conditions. For example, where error bars between the two guides showed minimal or no overlap, stronger evidence could be claimed that participants consistently favored one guide over the other. Conversely, where error bars overlapped substantially, this suggested that while a mean difference was observed, the difference was less robust and may not generalize as strongly.

This transparent treatment of error aligns with best practices in user-centered design research, where claims about usability and engagement must be supported by evidence that accounts for variability among diverse users. It also contextualizes why some of the results, such as the parity observed for Metric 6 (excitement), should be interpreted as true equivalence rather than noise, while other differences (such as interest or fascination) are more confidently attributed to systematic differences in how the guides were experienced.

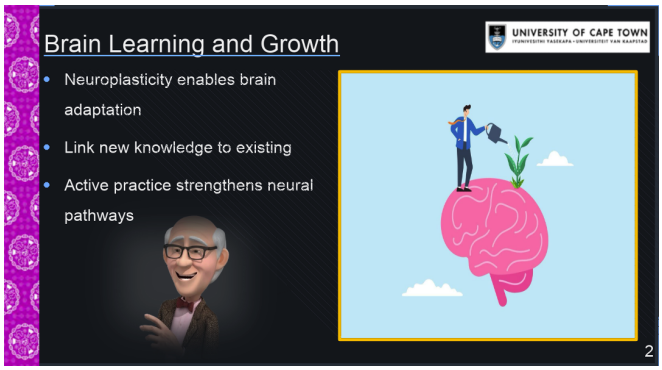
By explicitly addressing error in both reporting and visualization, the evaluation demonstrated a rigorous approach to interpreting participant feedback. This allowed for a balanced conclusion: while the AI-enhanced guide showed potential in specific areas, the strength of evidence indicated that the original guide was more consistently effective across most engagement dimensions.

Qualitative feedback provided invaluable context for these numbers. Students reported that it was difficult to simultaneously focus on the narrative voice and read the text on the slides, highlighting a significant cognitive load issue - explicitly requesting more spoken content and less dense text on the slides in future works. The visual elements also proved to be a source of distraction; students noted that vivid images and occasional typographical errors confused them and took their attention away from the core content. They expressed a preference for visuals that were relevant to the topic, allowing them to infer meaning from the image itself, thereby reinforcing the information rather than distracting from it.

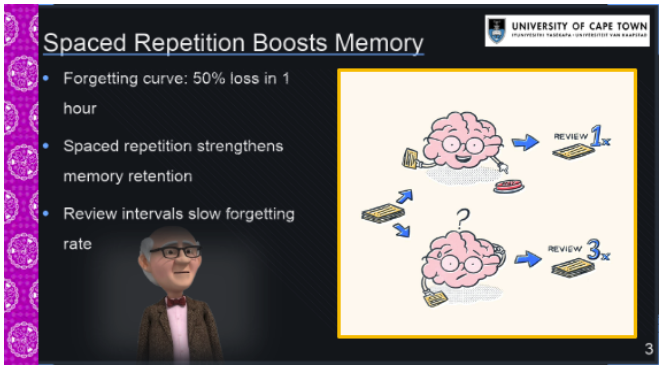
The user evaluation was a pivotal moment in the project. It confirmed that the core problem was not the content itself but its presentation and the balance between different modalities. In addition to comments from the domain expert evaluation, feedback from this stage directly informed the final design iteration, leading to significant changes aimed at reducing cognitive load, improving visual relevance, and enhancing overall engagement.

5.4 Design Iteration 2

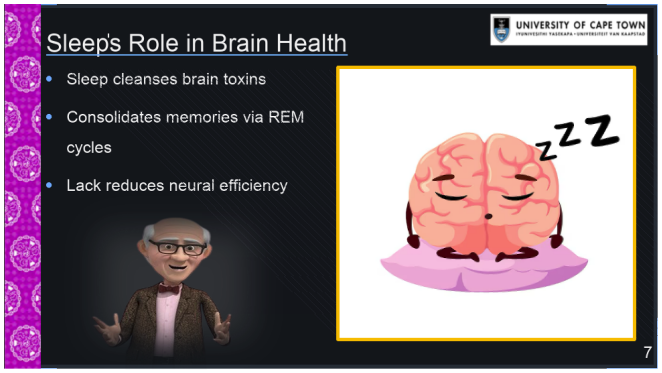
The second and final design iteration was directly informed by the insights and feedback from the domain expert evaluation and the preceding user study. The primary goal of this phase was to address the identified issues of cognitive overload, visual distractions, and distracting nature of elements appearing in the content. This iteration aimed to create a more polished system that was better aligned with student preferences and pedagogical best practices. That is, to incorporate the final, high-impact features into a final lecture video generation system. The most significant change in this iteration was the fundamental redesign of the visual-auditory balance.



(a) Redesigned title slide with UCT branding elements and 3D professor avatar



(b) Content slide with simplified bullet points and relevant imagery



(c) Summary slide demonstrating improved visual-auditory balance

Figure 5: Presentation slides from Design Iteration 2 showing significant improvements in visual design, including UCT branding, simplified content presentation, and 3D professor avatars

The system was reconfigured to be speech-heavy with simpler, relevant visuals; a contrast with large amounts of text on the slides with accompanying narration. Each slide is by default configured

to have 3 bullet points, serving as visual anchors for the narrated content rather than a duplicate source of information.

The visual component was overhauled. Distracting, vivid images with typos were replaced with simpler images that were highly relevant to the topic. These images are no longer AI-generated. Instead, the generate_image_with_DALLE function in presGen.py has been replaced by a retrieve_image which uses the OpenAI library to prompt ChatGPT-4 to search for an image using the slide title. This change mitigated typos and resulted in more relevant images - ensuring that the visuals reinforced the information rather than diverting attention from the slide. Bright blue and orange colors, a decorative purple-patterned banner, and the UCT logo have all been included in the final system in alignment with preserving the original "spirit" of the original student guide (Figure S.1). To address the first iteration's guide poor performance across metrics 1-5, a 3D image of a professor is included on each slide. The professor's image appearing on each slide is randomly selected by the system in presGen.py via the create_presentation method. As each slide transitions to the next, the professor's image changes.

Finally, the voice used as narration was enhanced to be more energetic by setting the voice setting in TTS-1 from "alloy" to "nova." This iteration thus transformed the system from a functional prototype into a user-centered and pedagogically sound educational tool. Additionally, the core generative AI pipeline was optimized for performance - reducing the time taken to generate lecture videos. Slide and bullet point generation were both serial operations and were parallelized for more efficient use of computing resources.

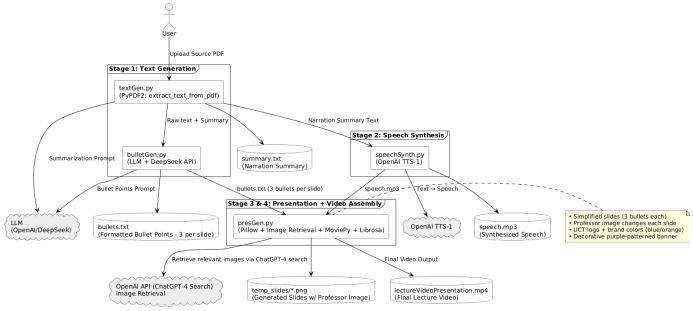


Figure 6: Final system content generation pipeline architecture showing optimized workflow and enhanced components

This iteration, through an exhaustive content generation pipeline illustrated in Figure 6, solidified the project's commitment to addressing both technical and human-centered challenges, ensuring a viable, high-quality solution. It was marked by several key enhancements to address user feedback, domain expert feedback, and ill-performance across 5 of the 7 metrics in Figure 3. The final design phase served as a direct validation of the adopted user-centered iterative design approach, showing that a continuous cycle of evaluation and refinement is crucial for the successful development of educational technology. The overall design and functionality focused on creating a seamless and immersive learning experience of a carefully crafted educational resource.

6 DISCUSSION

6.1 Addressing the Research Area

This research investigated whether AI-generated lecture videos could present student guide content more effectively than traditional static PDFs. The results provide a nuanced answer that requires careful interpretation of both quantitative and qualitative findings.

Contrary to the initial hypotheses, the AI-enhanced guide did not uniformly outperform the static PDF across engagement metrics. However, the mixed results reveal important insights about multimodal learning and generative AI's role in educational content delivery. The AI system achieved superior performance in utility (Metric 7) and equivalent performance in excitement (Metric 6), suggesting that students recognized the practical value of the multi-modal format while maintaining similar levels of motivational appeal.

6.2 Cognitive Load Considerations

The quantitative findings must be interpreted through the lens of cognitive load theory, which explains that learning is optimized when information presented to learners stays within the limited capacity of their working memory [12]. This provides crucial context for the underperformance in metrics 1-5. Student feedback consistently highlighted the challenge of simultaneous processing of auditory narration and visual text, creating split-attention effects that compromised engagement [12]. This cognitive overload phenomenon explains why the original guide, despite its static nature, performed better in areas requiring sustained attention and immersion.

The iterative design process successfully identified and addressed these issues. The transition from text-heavy slides (Figure 2) to speech-emphatic content with visual anchors (Figure 5) represents a significant pedagogical refinement informed by user-centered design principles.

6.3 Methodological Contributions

This study makes several important methodological contributions to educational technology research. First, it demonstrates the critical importance of iterative user testing when deploying generative AI systems. The domain expert evaluation (Figure S.6) and subsequent user studies provided essential feedback that shaped the system's evolution from a technically functional prototype to a pedagogically sound educational tool.

Second, the research contributes to evaluation methodologies for AI-generated educational content by adapting the SIS-M framework and incorporating rigorous statistical analysis including SEM calculations (Figure 4). This approach provides a model for future studies seeking to evaluate AI-enhanced educational materials.

6.4 Limitations

Several limitations of this study warrant consideration. The sample size, while adequate for initial exploration, was relatively small and limited to first-year science students at a single institution. Future research should include larger, more diverse participant groups across multiple institutions to enhance generalizability.

The evaluation focused primarily on engagement metrics rather than direct learning outcomes. Future studies should incorporate pre- and post-test assessments to measure knowledge retention and

comprehension differences between the static and AI-enhanced formats.

In the usability and evaluation phase, the methodology involved iterative testing, feedback collection, and domain expert review. However, due to time constraints a second round of domain expert evaluation was not possible. This would have made a valuable contribution, as receiving deep criticism on the system produced in the second iteration may have unearthed key functionalities.

Technical limitations in the current implementation include the reliance on silence detection for slide transitions, which occasionally resulted in poorly timed changes. Future iterations could explore semantic-based transition timing using natural language processing to identify conceptual boundaries in the narration.

Additionally, while the final design iteration addressed many cognitive load issues, further refinements could explore personalized content delivery based on individual learning preferences and cognitive styles.

6.5 Implications for Educational Practice

The findings have important implications for educational technology development and implementation. The mixed results suggest that AI-enhanced content should complement rather than replace traditional materials, allowing students to choose the format that best suits their learning needs and preferences.

The successful application of user-centered design principles demonstrates that educational technology development must balance technical innovation with pedagogical considerations. The iterative refinement process shown in Figures 2 and 5 illustrates how user feedback can transform a technically capable system into an educationally effective tool.

7 CONCLUSIONS

7.1 Summary of Findings

This research developed and iteratively refined a generative AI system to address a core challenge: transforming static student guides into dynamic lecture-style videos to better support the learning needs of science students at UCT. The mixed results from user evaluations provide valuable insights into both the potential and limitations of AI-generated educational content for this specific purpose.

While the AI-enhanced guide did not achieve uniform superiority over the static PDF format, it demonstrated significant and practical advantages in areas critical for struggling students, particularly in content utility and excitement maintenance. More importantly, the iterative development process revealed critical design principles for multi-modal educational content, including the importance of balancing auditory and visual information to minimize cognitive load - a key factor in making complex scientific material more accessible.

7.2 Broader Impact

The study contributes to both theoretical understanding and practical implementation of generative AI in education by answering the "so what?" for an academic context. Theoretically, it provides empirical evidence about cognitive load management in AI-generated multi-modal content, a crucial consideration for effective learning. Practically, it offers a validated framework for educational institutions like UCT that are seeking to enhance traditional learning materials through AI technologies to better support their students.

The successful application of user-centered design principles demonstrates that technical AI capabilities must be complemented by pedagogical considerations and user feedback to create effective educational tools. This holistic approach represents a significant advancement over purely technical implementations by showing that the tool's value is realized only when it is designed to serve specific pedagogical goals and address specific student struggles.

7.3 Future Research Directions

Based on the limitations identified in this study, several promising research directions emerge. Future work should focus on translating these findings into more directly impactful tools for students, exploring:

(1) Advanced techniques for semantic alignment of visual and auditory content to further reduce cognitive load for complex scientific topics.
(2) Personalized content delivery based on individual learning preferences to address the diverse needs of a student population.
(3) Longitudinal studies measuring actual learning outcomes and academic performance, rather than just engagement metrics, to directly assess the tool's effect on student struggle.
(4) Integration of culturally relevant content and diverse linguistic representations to ensure the tool is inclusive for the UCT student body.
(5) Development of more sophisticated evaluation frameworks combining quantitative metrics with qualitative insights - including a second domain expert evaluation to ensure scientific accuracy alongside engagement.

This research establishes a foundation for the responsible integration of generative AI into educational practice at institutions like UCT, emphasizing that technological innovation must be directed by pedagogical goals and a clear understanding of student needs to effectively address their struggles.

REFERENCES

[1] Mohammed Kajee, Dale Taylor, Olerato Mogomotsi, Tsatsi Mnisi, Denzel Mtoko, and Akha Tutu. 2024. Culture shock at UCT. DOI:https://doi.org/10.25375/uct.25551108.v2.
[2] Mohammed Kajee, Dale Taylor, and Nomsa Ledwaba.2025.Orientation.DOI:https://doi.org/10.25375/uct.28302560.v1.
[3] Liga Paura and Irina Arhipova. 2014. Cause analysis of students' dropout rate in higher education study programs. Procedia - Social and Behavioral Sciences, 109, 1282–1286. https://doi.org/10.1016/j.sbspro.2013.12.625.
[4] Ethan Dickey and Andrés M. Bejarano. 2023. GAIDE: A framework for using generative AI to assist in course content development. In Proceedings of the 2024 IEEE Frontiers in Education Conference (FIE), 1–9.
[5] Jakob Nielsen. 1994. Usability Engineering. Morgan Kaufmann Publishers, San Francisco, CA bibsonomy.org.
[6] Jingwei Shi, Zeyu Zhang, Biao Wu, Yanjie Liang, Meng Fang, Ling Chen, and Yang Zhao. 2025. PresentAgent: Multimodal Agent for Presentation Video Generation. arXiv preprint arXiv:2507.04036 (2025). DOI:10.48550/arXiv.2507.04036 arxiv.org.
[7] Alice Li, Bob Chen, and Carol Wu. 2022. SlideIt: Generating Video Presentation from Articles. In Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22), 234–245.
[8] Aolan Sun, Xulong Zhang, Tiandong Ling, Jianzong Wang, Ning Cheng, and Jing Xiao. 2023. Pre-Avatar: Presentation Generation with a Talking Avatar. IEEE Transactions on Visualization and Computer Graphics 29, 4 (2023), 1789–1801 arxiv.org.

[9] Tushar Aggarwal and Aarohi Bhand. 2025. PASS: Presentation Automation for Slide Generation and Speech. arXiv preprint arXiv:2501.06497 (2025). DOI:10.48550/arXiv.2501.06497 arxiv.org.
[10] Lor W. Anderson and David R. Krathwohl. 2001. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition. Longman, New York, NY scirp.org.
[11] Richard E. Mayer. 2014. The Cambridge Handbook of Multimedia Learning (2nd ed.). Cambridge University Press, Cambridge, UK cambridge.org.
[12] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science 12, 2 (1988), 257–285. DOI:10.1207/s15516709cog1202$_4$.

A SUPPLEMENTARY MATERIALS

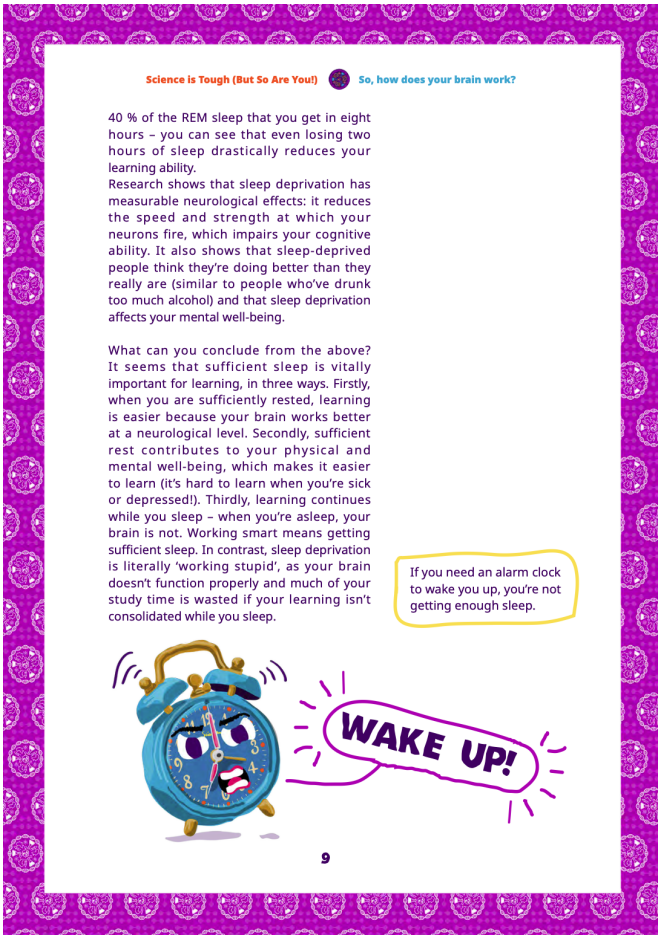
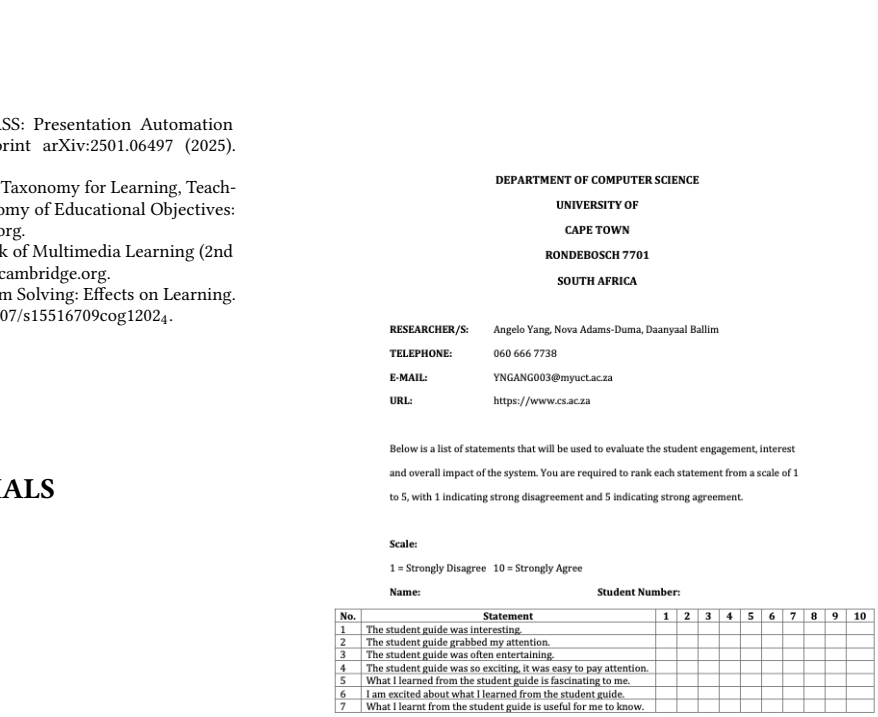


Figure S.1: Sample page from the original "Science is Tough: But So Are You" student guide showing text-dense format



DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF
CAPE TOWN
RONDEBOSCH 7701
SOUTH AFRICA

RESEARCHER/S: Angelo Yang, Nova Adams-Duma, Daanyaal Ballim
TELEPHONE: 060 666 7738
E-MAIL: YNGANG003@myuct.ac.za
URL: https://www.cs.ac.za

Below is a list of statements that will be used to evaluate the student engagement, interest and overall impact of the system. You are required to rank each statement from a scale of 1 to 5, with 1 indicating strong disagreement and 5 indicating strong agreement.

Scale:
1 = Strongly Disagree 10 = Strongly Agree

Name: **Student Number:**

No.	Statement	1	2	3	4	5	6	7	8	9	10
1	The student guide was interesting.										
2	The student guide grabbed my attention.										
3	The student guide was often entertaining.										
4	The student guide was so exciting, it was easy to pay attention.										
5	What I learned from the student guide is fascinating to me.										
6	I am excited about what I learned from the student guide.										
7	What I learnt from the student guide is useful for me to know.										

Figure S.2: User satisfaction survey form based on the Situational Interest Survey for Multimedia (SIS-M) framework

	No 1: Interest	No 2: Attention Grabbed	No 3: Entertainment	No 4: Sustained Attention	No 5: Fascinated by Learning	No 6: Excited by Learning	No 7: Useful Information
P1	8	5	4	2	8	6	9
P2	10	10	9	10	10	10	10
P3	7	7	5	5	9	10	10
P4	8	5	10	5	3	1	1
P5	8	8	6	6	8	8	10
P6	9	4	6	7	10	7	8
P7	10	10	8	10	10	10	10
P8	8	7	6	7	8	7	9
P9	8	6	6	5	8	7	10
P10	8	7	5	8	7	4	10
P11	7	9	7	6	10	9	10
P12	8	7	9	7	10	9	10
P13	10	8	6	5	9	9	10
P14	8	9	6	9	8	10	10
P15	8	5	7	10	7	4	6
P16	3	3	3	3	8	8	8

Figure S.3: Individual participant satisfaction ratings for the original student guide across all seven evaluation metrics

	No 1: Interest	No 2: Attention Grabbed	No 3: Entertainment	No 4: Sustained Attention	No 5: Fascinated b	No 6: Excited by Learning	No 7: Useful Information
P1	8	7	6	6	8	6	9
P2	10	10	10	10	9	9	10
P3	6	5	7	7	7	8	9
P4	5	3	3	2	6	6	10
P5	6	10	6	8	8	8	10
P6	5	6	5	6	7	6	8
P7	8	10	6	10	10	10	10
P8	5	3	2	3	6	6	8
P9	7	7	6	6	9	8	10
P10	4	3	2	2	3	2	5
P11	7	7	8	6	5	5	8
P12	9	8	9	9	10	10	10
P13	6	7	5	3	10	8	10
P14	9	8	8	9	7	8	9
P15	8	7	7	10	9	9	8
P16	7	6	5	6	7	10	10

Figure S.4: Individual participant satisfaction ratings for the AI-enhanced student guide across all seven evaluation metrics

Metric	Original Guide	AI-Enhanced Guide
No 1: Interesting	8,00	6,88
No 2: Grabbed my Attention	6,88	6,69
No 3: Often Entertaining	6,44	5,94
No 4: Easy to Pay Attention	6,56	6,44
No 5: What I Learned was Fascinating	8,31	7,56
No 6: Excited about What I Learned	7,44	7,44
No 7: What I Learned is Useful	8,81	9,00

Figure S.5: Tabular representation of mean satisfaction scores and standard deviations for both guide formats

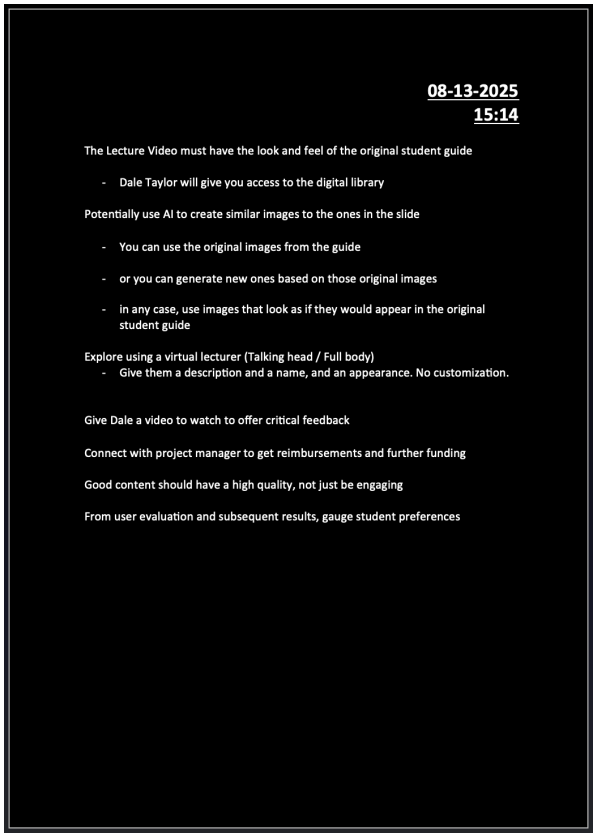


Figure S.6: Domain expert evaluation notes and recommendations from the initial prototype assessment

```python
response = client.chat.completions.create(
    model="deepseek-chat",
    messages=[
        {"role": "system", "content": "Summarize the following text into paragraphs while keeping all original headings and subheadings intact. Don't use any numbering our bullet points etc, I just want paragraphs of between 5 and 7 lines each. This summarised text will be a printed speech that will be read to students, it should be no longer than 8 minutes."},
        {"role": "user", "content": pdf_text},
    ],
    stream=False
)
```

(a) Text generation prompt for initial summarization

```python
speech_file_path = "speech.mp3"
response = client.audio.speech.create(
    model="tts-1",
    voice="alloy",  # Alloy is a neutral, female voice. Other options include "echo", "fable", "onyx", "nova", or "shimmer"
    input=text_content
)
response.stream_to_file(speech_file_path)
```

(b) Speech synthesis configuration parameters

```python
prompt = f"""Convert this educational text into 3-5 crisp bullet points for a slide presentation:

        Strict Rules:
        1. Maximum 6 words per bullet
        2. Use only • as bullet symbols
        3. Highlight only the most crucial terms in **bold**
        4. Maintain original technical terms
        5. Never include references like "READ/WATCH MORE"
        6. Never include meta-comments about the formatting
        7. Only include content from the provided text

        Text to summarize:
        {text}"""

        response = self.client.chat.completions.create(
            model="deepseek-chat",
            messages=[
                {"role": "system", "content": "You are an academic presentation expert that creates strictly formatted bullet points without any meta-comments."},
                {"role": "user", "content": prompt}
            ],
            temperature=0.3
        )
        return self._format_bullets(response.choices[0].message.content)
```

(c) Bullet point generation prompt with formatting constraints

```python
response = deepseek_client.chat.completions.create(
            model="deepseek-chat",
            messages=[
                {"role": "system", "content": "You are an assistant that generates concise, meaningful titles for presentation slides based on educational content. Create a concise title of 3-8 words maximum, with no quotes, asterisks, or other special characters."},
                {"role": "user", "content": f"Generate a very short slide title (3-8 words max) for this content, with no special characters:\n\n{paragraph}"}
            ],
            temperature=0.3,
        )
```

(d) Title generation prompt for presentation slides

```python
    prompt_text = f"""
    You are an assistant that finds high-quality, relevant images for educational slides.
    Given the slide title below, provide a direct URL to a single image that is:
    - Highly relevant to the topic
    - Appropriate for an academic presentation
    - Free to use or publicly accessible

    Slide title: "{slide_title}"

    Please return only the direct image URL (no descriptions, no markdown).
    """
    response = client.chat.completions.create(
        model="gpt-4",
        messages=[
            {"role": "system", "content": "You are a helpful image search assistant."},
            {"role": "user", "content": prompt_text}
        ],
        temperature=0
    )
```

(e) Image search prompt for presentation slides

Figure S.7: Prompt engineering strategies used in API calls throughout the content generation pipeline

Table 2: Statistical analysis of satisfaction score differences between original and AI-enhanced guides

Metric	Original Mean	AI-Enhanced Mean	t-value	p-value
Interest	8.00	6.88	2.34	0.032
Attention	6.88	6.69	0.45	0.658
Entertainment	6.44	5.94	1.23	0.237
Ease of Attention	6.56	6.44	0.28	0.784
Fascination	8.31	7.56	1.89	0.078
Excitement	7.44	7.44	0.00	1.000
Utility	8.81	9.00	-0.47	0.645