Generative AI Applied to Student Guide

Angelo Yang
University of Cape Town
Cape Town, South Africa
YNGANG003@myuct.ac.za

ABSTRACT

The Generative AI Applied to Student Guide (GAISG) is a web-based application developed to address the low engagement rates of the University of Cape Town "Science is Tough: But So Are You" student guide. The student guide is targeted towards first-year Science students, however due to its lengthy and static nature it fails to engage students. As such, this paper focuses on improving student engagement rates through the generation of personalized podcasts using generative AI. To moderate the generated content, the system integrates the GAIDE moderation framework, which ensures podcasts remain aligned with the original guide. In addition, the paper explores what AI agent personalities and demographics students prefer by enabling students to create their own AI hosts. Subsequently, a user study with sixteen first-year students evaluated engagement levels and explored preferences for AI host demographics and personalities. Results show consistently higher average engagement scores for podcasts compared to the original guide along with a student preference for host characteristics that align with their own.

CCS Concepts

• Generative AI

1 INTRODUCTION

The University of Cape Town (UCT) "Science is Tough: But So Are You" student guide is a supportive web-based guide that helps first-year science students navigate through the challenges of a BSc degree. The student guide can be accessed via Amathuba – UCT's primary teaching platform. This guide is a website with links to 13 PDF documents – each detailing methods that first-year science students are advised to adopt for their transition into university life. This student guide comprehensively addresses many struggles students experience in their first year at university. However, the static, lengthy and non-interactive nature of its content fails to effectively engage readers. Evidence of this can be observed in the most popular guide, *Culture Shock at UCT* having only 606 views and 102 downloads [1], whereas the least popular, *Orientation* has just 24 views and 19 downloads [2].

This is a critical failure, the student guide is typically used to help first-year Science faculty students transition from high school into university culture. As a result, students are inadequately equipped for university environments, time management demands and academic workload. Given these limitations, there is a clear need for more interactive and engaging formats. Generative AI podcasts offer one such solution, combining accessibility with personalization to better support first-year students.

1.1 Problem Statement

The central issue is that the current "Science is Tough: But So Are You" student guide undermines its purpose in addressing students' challenges in a timely, effective, and engaging manner. This greatly affects the degree to which first-year science students at UCT find such a guide appealing, as information is densely populated without engaging features that captivate the readers' attention. Given that this resource is aimed at helping these students transition into university from high school, such students may instead become inadequately equipped due to lack of engagement with the student guide. Many students find the first year of university the most challenging. A study by Paura et al. shows that up to 23.2% and 11.2% of students drop out in the first six and twelve months of university respectively [3]. Among first-year students, 50% study for tests at the last minute, 54% struggle transitioning from high school to university, 33% skip classes, and 25% are not a part of peer mentor groups, according to another study by Barlow-Jones [4].

Furthermore, a recent study by Eliza Lily Goode et al. detailing the impact of interactive learning models and synchronous class attendance on student achievement made a key conclusion. The study found that - through qualitative data - students associate interactive, media rich, flexible, and responsive online modules with higher levels of engagement and deeper learning [5]. As such, this project aims to raise student engagement by integrating generative AI audio technologies into an interactive student guide.

1.2 Research Area

This paper will focus on two interconnected research areas. Firstly, the application of generative AI in creating personalized podcasts that are accessible and engaging. The central research question is "To what extent do first-year science students perceive the AI-generated podcast as more engaging than traditional student guides?". This will be investigated through the implementation of AI-generated podcasts to which first-year students will evaluate its effectiveness. The paper hypothesizes that students will report higher levels of engagement when using the proposed student guide podcasts compared to the existing PDF-based student guide.

Secondly, the paper will explore the research question "What AI agent personalities and demographics do first-year science students prefer?". This research question will be answered through the personalization aspect of the AI-generated podcasts. The personalization features will enable students to customize their podcast hosts by selecting attributes such as age, gender, accent, and speaking styles. The paper hypothesizes that students will prefer AI host demographics and personalities that align with their own characteristics.

Together, these questions will contribute to the broader field of generative AI in education, examining both the effectiveness of AI-generated podcasts as a learning tool and the qualities that students value in their AI hosts, thereby addressing the limitations of UCT's current student guide.

1.3 Functional and Non-Functional Requirements

The podcast system adopted the Agile design process, with the first stage focusing on gathering requirements. This process involved consultations with key stakeholders, the supervisor Gary Stewart and Dale Taylor, the editor of the "Science is Tough: But So Are You" student guides. Four crucial requirements were identified from the series of meetings with the stakeholders.

First, the system's user interfaces should be intuitive and straightforward to navigate, adhering to at least 9 of Jakob Nielsen's 10 Usability Heuristics [6]. This ensures that complex or confusing user interfaces do not prevent students from fully engaging with the podcasts.

Second, given that students will have direct access to the AI-generated podcasts, a moderation framework must validate all outputs. The full GAIDE framework should be integrated to safeguard against inappropriate, misleading, or misaligned content.

Third, to provide a true personalized experience, the system should offer a sufficiently large voice library—at least 60 distinct AI voices—so that students can diversify their podcast creations and select hosts that reflect their preferences.

Lastly, the podcast generation process should be responsive to maintain student engagement. Excluding external API requests, the system's internal generation time should not exceed 40% of the total podcast duration.

2 RELATED WORKS

2.1 Cinematic Clinical Narratives

This study by Bland et al aimed to enhance the teaching of clinical pharmacology in medical school by using a multimodal generative AI approach in creating engaging, cinematic clinical narratives (CNNs). The study used the following generative AI tools for the creation of CNNs, ChatGPT (GPT-4), Leonardo.ai and Stable Diffusion, ElevenLabs and Suno to create "Shattered Slippers", an engaging, interactive multimedia experience [7].

Each CNN consisted of four components: plot, images, narration, and theme song. The plot was developed by reimagining first-year clinical cases, with ChatGPT generating fictional storylines that were refined to maintain educational objectives while adding thematic resonance and real-world connections [7]. The image component used Leonardo.ai, built on Stable Diffusion [8], to create cinematic visuals of clinical scenes, incorporating popular culture and celebrities to increase relatability. Narration was produced by submitting the ChatGPT script to ElevenLabs [7][9] and paired with images such that narration automatically played with each scene, supporting multisensory learning [7]. For the theme song, ChatGPT-generated lyrics were input into Suno Chirp Bot to produce melody and vocals. Finally, all components were synchronized in a PowerPoint presentation.

The data collection phase of the "Shattered Slippers" CNN study involved forty first-year medical students, to which eighteen provided feedback. The feedback process utilized the Situational Interest Survey for Multimedia (SIS-M), a measurement tool developed to evaluate a learner's situational interest in multimedia based educational environments [10–12]. SIS-M was originally used to evaluate the effectiveness of multimedia in promoting engagement and motivation in higher education [10–11]. This survey tool was used to capture student feedback regarding their views and opinions on the "Shattered Slippers" CNN. The survey includes items to rank on a 1-5 scale), a question asking for preference of the CNN compared to traditional teaching methods, and lastly, an open-ended question asking, "Why do you think this is your preference." [7]. The specific survey items used in SIS-M are detailed in Appendix A [7].

The Cinematic Clinical Narratives (CNN) produced promising results in its implementation of generative AI into its course creation. The quantitative assessment of the "Shattered Slippers" CNN using the SIS-M survey indicated high levels in participants' interest with the generated content; with fourteen out of eighteen students indicating preference of CNNs over traditionally presented clinical cases [7]. The combination of text, image, audio and music created an immersive learning experience that captured students' attention, catered to different learning styles and increased engagement with the course content. Additionally, real clinical cases were used in generating the script which ensured thematic relevance and alignment with learning objectives.

However, there were noticeable drops in students' excitement about what they learned, their interest in the information presented in the CNN, and whether students felt that the content was important to them [7]. This can be attributed to the limitations of the CNN. As stated in the study, the incorporation of popular culture, specifically cultural references and celebrities, may not resonate with all learners [7]. This may result in further detachment of students from the course content as they feel underrepresented. Lastly, a formal validation framework was not implemented, which could lead to unreliable, irrelevant and misaligned content.

The paper draws four major inspirations from the study by Bland et al. Firstly, the generative AI pipeline used in the study to create the four components can also be applied to the proposed podcast system [7]. The podcast implementation would also consist of a script, audio, music and a final compilation stage. As such, the same modular approach will be applied, such that each component can be generated and refined individually, ensuring both engagement and alignment with educational objectives. Secondly, the paper will utilize the SIS-M survey presented in the CNN study. The survey is a validated instrument that has been previously applied in related contexts and aligns closely with the research objectives of this paper [7]. In addition, the paper will adapt the idea of incorporating popular culture and celebrities to increase relatability. This will take the form of a descriptive voice creation feature where students will be able to create their own host voice. Subsequently increasing interactivity and engagement with the student guide while preventing students from feeling underrepresented with the default voice options provided. Lastly, the use of a theme song will also be applied to the podcast system to create a feeling of immersion and enhance the overall learning atmosphere [7].

2.2 Personalized Programming Exercises

Logacheva et al. explored the use of generative AI, specifically large language models, to contextually personalize programming problems. The aim was to evaluate whether such personalization techniques could enhance student engagement by creating exercises suited to each student's interests and background [13]. Using ChatGPT (GPT-4), the study generated Dart programming

problems that followed standard beginner course structures but were enriched with real-world themes such as sports, music, and popular culture to increase relevance. Earlier research supported this approach, showing the motivational benefits of contextualized programming tasks, including improved retention (Guzdial) and reduced algebraic errors (Leinonen et al.) [14][15]. Following this, exercises were produced through structured prompt templates that paired programming concepts such as loops, conditionals, and strings with themes such as cooking, gaming, or sports.

The majority of generated problems were rated as clear and conceptually aligned, though some were judged too easy or only loosely related to their themes [13]. These findings reflect limitations observed in other work, such as del Carpio Gutierrez et al., who showed that while ChatGPT can generate high-quality exercises, human review is often necessary to ensure contextual depth and educational accuracy [16].

This paper draws one major inspiration from Logacheva et al. in its use of personalization to enhance student engagement. In addition, similar to Cinematic Clinical Narratives, the absence of a formal moderation framework resulted in partially relevant themes and misalignment with the intended learning objectives. To address this, the proposed podcast system will integrate the Generative AI for Instructional Development and Education framework, which is reviewed in the next related works.

2.3 Generative AI for Instructional Development and Education

GAIDE is a generative AI content generation framework detailing a sequence of steps in integrating generative AI into collegiate level course content development [17]. It outlines a six-step process that guides educators through setting up their learning goals, first draft generation, macro refining, micro refining, maintaining contextual integrity during refinement and consolidating generated content. The framework's general applicability aligns with the goals of the proposed student guide podcast and will form the backbone of the AI-generated content moderation and refinement process [17].

2.3.1 Setup

Educators should first define specific learning goals to be achieved. These goals should guide the narrative presented to the generative AI and serve as a constant reminder of the intended direction. The goals should also be specific in the measurable outcomes that should be achieved. Educators should then set up the context, this involves telling the model how to act, supplying student demographics, skill sets and background to tailor the content to their intended audience. Lastly, the learning objectives should be generated and their alignment with course outcomes ensured such that the generated content remains highly relevant and engaging for students [17].

2.3.2 First Draft Generation

In this study, Dickey et al. classifies course content into two forms, lecture-style and problem creation. A rough draft of a lecture-style course content involves generating an outline that is accurate at a high-level view, while the latter involves generating a list of potential problems to which a subset will be selected for iterative refinement [17].

2.3.3 Macro Refinement

The focus of this step is to refine the generated course content to a holistic level. This involves refining entire sections to ensure alignment with the established expectations, goals, learning objectives, and context [17]. For lecture-style course contents, the outline should be iteratively refined with the following factors taken into consideration, the duration of the lecture, associated tasks, pre- and post-lecture activities, subtopics and specific activities [17]. For the list of problems generated, if they lack alignment with the context and intended goals of the course, feedback should be provided to the generative AI, emphasizing the desired attributes.

2.3.4 Micro Refinement

Before moving onto this step, educators should be satisfied with the overall structure and alignment of the drafted content. From here, each section, subsection or question should be examined and refined while informing the generative AI on the focus of the section. Educators should prompt the AI to refine wording for improved clarity, adjust complexity to match student skill sets, or rephrase questions to promote critical thinking. In addition, educators should validate AI-generated exercises by requesting example answers from the model to ensure accuracy and alignment with intended learning outcomes [17].

2.3.5 Maintaining Contextual Integrity During Refinement

During iterative refinement, educators should ensure that context blending and loss of focus by the generative AI does not occur [17]. To mitigate context blending, educators should explicitly inform the model of transitions between course content sections and consistently reinforce the learning objectives, context and focus of the section. To mitigate loss of focus, new sessions may be started, and the context reintroduced to ensure that the generative AI model focuses on the current task at hand

2.3.6 Consolidating Generated Content

Once the learning objectives, course outlines and assessments are established, educators should compile the generated content into a cohesive and well-structured course, making final refinements where necessary.

2.3.7 Applications

As seen in the CNN study by Bland et al., a formal AI moderation framework was not implemented. This could lead to the generation of unreliable, irrelevant and misaligned content. In addition, due to the sensitive nature of presenting guidance material to first-year students, the podcasts will also need to be moderated in their cultural sensitivity, appropriateness and alignment with the original student guide.

To this end, the podcast system will integrate two adapted versions of the GAIDE framework – GAIDE Manual and Automatic – to guide a structured and iterative refinement process for the podcast system.

3 DESIGN

The design process for the podcast system followed the Agile design process, and as such, two iterations are presented. The first iteration was evaluated by first-year science students, with the results detailed and discussed in the later sections of the paper. In addition, an expert evaluator provided critical feedback on the

initial iteration. The second iteration was built upon the feedback and newly identified requirements from the first student evaluations and expert review. The changes and refinements of the final iteration are presented in the discussion of results. While this final iteration was not evaluated by students due to time constraints, it underwent a final assessment by Dale Taylor.

Following the requirements gathering stage presented in Section 1.4, these requirements were analyzed and informed the system's design choices as part of an Agile software development cycle. Usability is addressed through Jakob Nielsen's heuristics in both Creational and Premade modes. Moderation is enforced by integrating the GAIDE framework, with GAIDE Automatic for students and GAIDE Manual for educators. Personalization is enabled through customizable voice options in the Creational mode, and responsiveness is ensured via optimizations such as parallelized text-to-speech (TTS) requests and cloud storage integration. The following subsections elaborate on each of these design decisions in detail.

3.1 Modes

The podcast system features two podcast generation modes. The Creational Mode is designed for students to generate podcasts in real time, fostering personalization and active engagement with the student guide content. By contrast, the Premade Mode is intended for educators and student guide authors, such as Dale Taylor, to efficiently create moderated podcasts which are uploaded and accessible to students. Both modes provide customization through a set of default podcast options, with the specific parameters outlined in Table 2 below. While the first three parameters can be found in common podcast configurations, the introductory theme music provides an additional layer of immersion.

Table 2: Default Podcast Parameters

Parameter	Description
Student Guide	The student guide content used to generate the podcast script, in the form of extracted text from the PDF.
Number of Hosts	The number of hosts presented in the podcast. (Minimum: 1, Maximum: 2)
Podcast Length	The length of the podcast in minutes. (Minimum: 1, Maximum: 10)
Podcast Intro	Whether the introductory theme music is included in the podcast.

3.1.1 Creational

Given that the Creational mode aims to foster personalization and interaction, a diverse selection of options is provided to enable students to tailor the podcast experience to their individual preferences. The default podcast parameters are outlined in Table 2, with the additional options available to students presented in Table 3.

Table 3: Additional Creational Podcast Parameters

Host Parameters	Description
Host (1/2): Age	The perceived age range of the podcast host's voice. (Young-Adult, Midlife Voices, Seasoned Senior)
Host (1/2): Accent	Specifies the regional or cultural accent applied to the host's speech (American, African American, British, Indian, South African)
Host (1/2): Gender	Sets the gender identity expressed through the host's voice (Male, Female)
Host (1/2): Tone	Controls the emotional delivery of the hosts. (Relaxed, Fun, Professional, Confident, Energetic)
Host (1/2): Style	Determines the host's speaking style. (Conversational, Educational, Entertaining)

The additional parameters provided in the Creational mode also serve to investigate the research question "What AI agent personalities and demographics do first-year science students prefer most?". While the diverse selection of options largely encapsulates the range of AI host personalities, additional parameters and choices are constrained by the range of voice options available through the selected text-to-speech API. This also enforces the requirement of at least 60 distinct AI voices with the host accent, gender and style combining to a total of 90 voices. The tone parameter is not included in the host voice selection, as incorporating it would increase the combinations to 360 and require an additional 270 manual selections from the text-to-speech voice library. To this end, the 90 featured voices fulfil the requirement of at least 60 distinct AI voices to support full personalization.

In terms of accuracy, the voices selected by students are constrained to and aligned with the parameter set available within the chosen text-to-speech API, ensuring consistency between user input and generated output. In addition, to ensure that generated podcasts remain appropriate and aligned with the student guide, GAIDE Automatic moderates content in the background, without requiring student intervention. The full design and implementation of GAIDE Automatic is presented in Section 3.2.2. Automatic and 4.4.2 Automatic respectively.

Thereafter, the student will simply create and view the podcast. The sequence of actions are detailed in Figure 1 below.

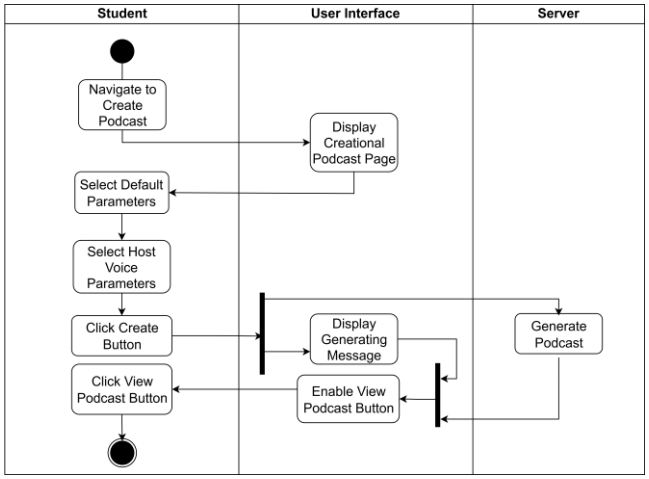


Figure 1: Creational Podcast Activity Diagram

3.1.2 Premade

Given that the Premade mode is intended for educators and student guide authors, efficiency, moderation and validation are the main objectives. Due to the lengthy process of moderating the script, the educator will only be presented with the default podcast parameters, as presented in Table 2, and premade voice options stored within the podcast system; of which a description of the voices will be provided. The available voice options will be sourced from the selected text-to-speech API's voice library, ensuring quality and alignment with its description.

While the Creational mode investigates both research questions, the Premade mode provides educators with an interface to create fully moderated podcasts. In addition, it investigates the feasibility and effectiveness of the GAIDE framework. To this end, an adapted version of the moderation framework will be integrated into the system. Educators will need to perform the six-step process [17] of manually reviewing the podcast structure and script before generating the podcast audio.

The activities performed by educators are outlined in Figure 2 below, while the full design and implementation of GAIDE Manual are presented in Section 3.2.1 Manual and 4.4.1 Manual respectively.

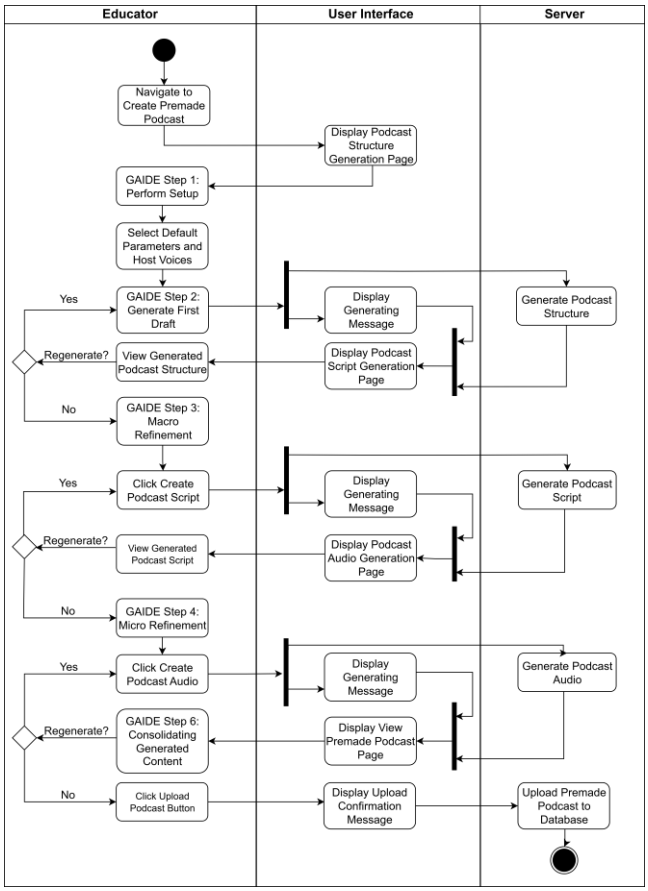


Figure 2: Premade Podcast Activity Diagram

Figure 2 above illustrates that the full GAIDE framework, except Step 5, has to be performed before the final audio is generated. Step 5 is automatically applied as will be explained in the next section.

3.2 GAIDE

3.2.1 Manual

The design for GAIDE Manual largely follows the framework by Dickey and Bejarano. The setup and first draft generation steps are combined into one, where the educator will provide the student guide content for the generation of the podcast. The learning goals and context are embedded into the prompt submitted to the script generator AI. This streamlines but preserves the setup step. Subsequently, the generation of the podcast structure fulfills Step 2. Thereafter, the structure in its summary and objectives are displayed. The macro refinement stage is then performed as the educator reviews and edits the podcast structure. If the structure is acceptable, the educator generates the first draft of the script and proceeds to Step 4, micro refinement of the script. If the structure is not acceptable in its entirety, the educator may regenerate the entire structure. Similar to the previous step, the educator is able to review the entire podcast script and perform micro refinements. If the script is acceptable, the educator proceeds to generating the audio. Similarly, they may also choose to regenerate the entire script. The fifth step in maintaining contextual integrity during iterative refinement is automatically enforced by the system. This is a result of all prompts reiterating the learning objectives and context. In addition, context blending [17] is mitigated by the modular pipeline approach where the structure and script requests are submitted separately. A loss of focus is also avoided by the fact that each request is submitted individually and treated as separate sessions [17]. After the educator generates the audio, they perform the final step in consolidating the generated content. This involves reviewing the audio, performing final refinements and attaching the title, author, description and theme song to the audio file. The educator may also choose to regenerate the entire podcast audio. In addition, a back button extends GAIDE's iterative process [17] by enabling educators to return to earlier stages when necessary.

3.2.2 Automatic

The design for GAIDE Automatic follows the framework to a lesser degree to ensure student usability, as manually performing the full moderation process would reduce efficiency and discourage engagement. Instead, moderation occurs automatically in the background. Setup and first draft generation are simplified into a single step where students select parameters and submit the preferred student guide for podcast creation. After the script is produced, it is resubmitted for automated evaluation against learning objectives, contextual integrity, and alignment. If the score falls below 90, the script is resubmitted, with the refinement process capped to three iterations to balance quality and responsiveness. Macro and micro refinements are thus hidden from the user, while Step 5 is enforced through repeated context embedded prompts, automated checks, and modular pipeline design. Finally, the consolidation step is handled by the system, presenting a moderated podcast to the student.

3.3 System Architecture

The system follows the layered architecture with four distinct layers: presentation, logic, service and data layer. The full system architecture is illustrated in Figure 3 as a package diagram, which highlights the four layers.

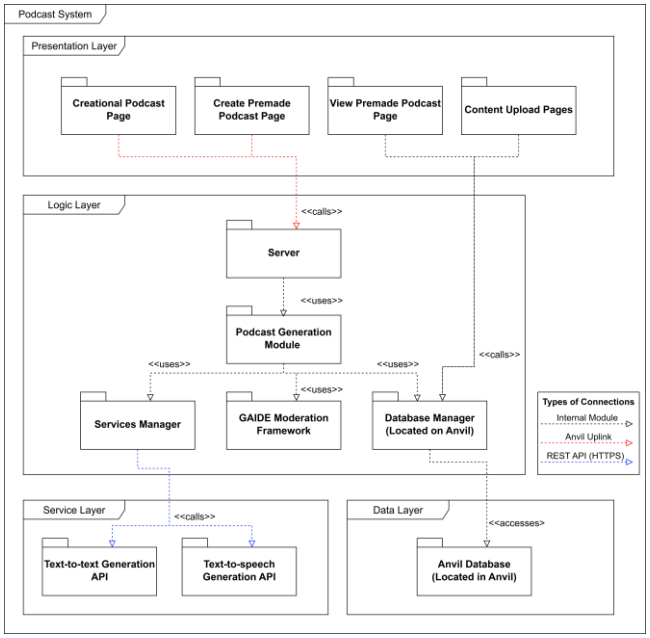


Figure 3: System Architecture Package Diagram.

Figure 3 presents the four layers of the podcast system; the View Premade Podcast Page module encapsulates all pages within its process. Similarly, student guide and voice uploading pages are grouped under the Content Upload Pages module. All presentation layer modules are hosted entirely within the Anvil cloud platform, a website development tool that offers client-server interfaces. On the other hand, all logic modules, except the Database Manager, run on an external server. The Database Manager itself will persist on Anvil's own server where it can directly communicate with the Anvil Database. Communication between the presentation and logic layer occurs over Anvil Uplink, Anvil's client-server interface. Lastly, all requests to generative-AI services utilize REST APIs (HTTPS).

This architecture offers three main advantages that align with the podcast system's requirements and goals. The first being the separation of concerns, core modules are decoupled from each other such that changes in one sub-system won't affect others. Secondly, maintainability and extensibility are enforced as new services can easily be swapped in without having to update data management and moderation modules. Lastly, performance is enhanced as service calls and audio generation processes live below the user interfaces; the system can scale the logic and service layers independently to handle peak student usage.

3.4 Technologies

The choice of text-to-speech generative AI technologies was informed by the requirements and objectives of the podcast system. As such, two hard requirements were identified. The first requirement needed the selected text-to-speech API to provide a large library of AI voices to enable students to personalize their podcasts. In addition, these voices had to be stable, natural and expressive to effectively engage students. The two hard requirements stated above were critical to answering both research questions. Soft requirements included ease of implementation, efficiency in response time to avoid disrupting student engagement, and affordability for long-term use in an educational context. To this end, Appendix B presents the preliminary evaluations of current text-to-speech technologies.

Initially, PlayAI's PlayHT suited the podcast system the most with its diverse and highly expressive voice library. Despite the platform's mediocre stability and affordability, it conveyed emotion far better than the other models. In addition, it featured in-text tags such as [laugh] and [energetic] which further boosted its authenticity and engagement. Lastly, its ease of implementation far surpassed the others as PlayHT's API allowed for multi-speaker requests, this allowed for two host scripts to be submitted simultaneously. However, midway through implementation, PlayHT was acquired by Meta, and future token purchases were subsequently discontinued. Subsequently, the system adopted ElevenLabs as its text-to-speech generator. Although it lacked the expressiveness of its predecessor, ElevenLabs offered superior stability and access to a wider range of voices.

In terms of the text-to-text generative AI, the podcast system required a platform capable of producing cohesive and contextually aligned scripts that maintained a natural conversational flow. As such, ChatGPT was selected, as it consistently generated content that aligned with the learning objectives of the student guide. This choice is further validated by related studies such as CNNs and Personalized Programming Exercises, both of which relied on ChatGPT.

Excluding core AI technologies, ElevenLabs Music was used to generate an introductory theme song. While Google Cloud Storage enabled educators to review student generated podcasts as well as the optimization of the generative AI pipeline, as will be discussed in Section 4.6 Optimization.

The full suite of technologies used in the podcast system are presented in Table 4 below.

Table 4: Technologies and Models Employed

Requirement	Platform	Model
Text-to-text	ChatGPT	GPT-4o
Text-to-speech	ElevenLabs	eleven_multilingual_v2
Introductory Music	ElevenLabs	Music Generation
Cloud Storage	Google Cloud	N/A

Lastly, for the development of the website frontend, Anvil was chosen due to its ease of drag-and-drop UI builder and integrated database support which streamlined the design. In addition, its template feature allowed one navigation bar to persist across all pages, enhancing usability. Prior experience with the platform also allowed for fast prototyping, which enabled multiple iterations to be produced. However, the client-server interface utilized Anvil Uplink which introduced high latency, this resulted in multiple sub-systems requiring optimizations.

3.5 Usability

Student and educator usability are essential objectives of the podcast system. Complex, unresponsive and unintuitive systems would hinder student interactivity and subsequently their engagement with the student guide. On the other hand, educators would waste valuable time navigating the system.

To mitigate these risks, the podcast system's user interface design follows Jakob Nielsen's 10 Usability Heuristics for User Interface Design [6]. These heuristics provide a well-established framework for creating intuitive, efficient, and user-friendly systems. By applying principles such as visibility of system status, consistency and standards, error prevention, and flexibility of use, the interface will reduce unnecessary complexity. The user interfaces and their adherence of the 10 heuristics will be presented in Section 4.8 User Interfaces.

In addition, to reduce the latency in generating and transmitting large audio files across different system layers, two different optimization strategies were implemented. The first of which parallelized the large number of text-to-speech requests sent to ElevenLabs. The second employed Google Cloud Storage buckets to temporarily store generated podcasts and allowed the server to return a single URL string to the Anvil frontend.

While the design choices address the four core requirements, several trade-offs were necessary. The Creational mode focuses on usability and engagement by simplifying GAIDE into an automated process, but this comes at the cost of fine-grained validation available in the Manual approach. On the other hand, the Premade mode enforces strict moderation, ensuring quality and alignment, but sacrifices the speed and personalization valued by students.

4 IMPLEMENTATION

The implementation continues the Agile software development process with descriptions of the core modules within the podcast system.

4.1 Generative AI Pipeline

The generative-AI pipeline serves as the backbone for the podcast system. The workflows for both modes are detailed through the Creational and Premade diagrams, shown in Figure 4 and Figure 5 respectively. These diagrams present the end-to-end pipeline and parameters required at each stage.

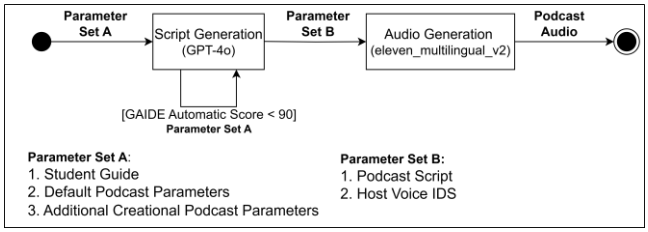


Figure 4: Creational Podcast Generative AI Pipeline.

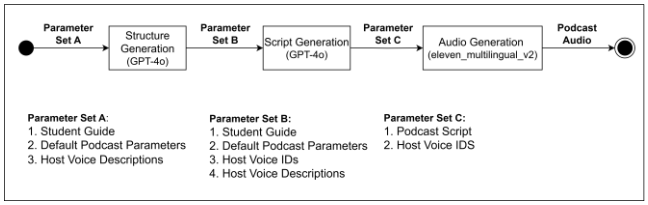


Figure 5: Premade Podcast Generative AI Pipeline

In Figure 4 and Figure 5, it can be observed that the student guide and host voice descriptions are resubmitted during the GAIDE refinement process. These iterative resubmissions ensure that both the content and host delivery remain aligned with the original learning objectives, while also reinforcing contextual integrity at each stage of the pipeline. Following this, the specific text generation implementations are presented in the next section.

4.2 Text Generation

The text-to-text generation process begins when the student or educator submits their request. Depending on whether a Creational or Premade podcast request is submitted, different structured prompts are used. The sections below will outline their structure and design.

4.2.1 Creational

The Creational podcast script structure requires both the Default Podcast Parameters and Additional Podcast Parameters, as detailed in Table 2 and Table 3 respectively. The student guide parameter guides ChatGPT in creating highly cohesive and relevant scripts while the length parameter determines the overall length of the podcast. The voice descriptions within the Additional Podcast Parameter influence host dialogue by embedding age, accent, gender, tone, and style into the script, ensuring it reflects the student's selected host personalities. The exact structured prompt implementation is detailed in Appendix C.

4.2.2 Premade

Similar to the student mode, the Premade podcast structure and script require the Default Podcast Parameters. In addition, it requires the voice descriptions attached to the selected voice. Together they define the overall structure and delivery of the podcast, ensuring consistency between the moderated script and the generated audio output. The exact structured prompt for the podcast structure is detailed in Appendix D and the script in Appendix E.

The voice descriptions embedded in both Creational and Premade podcasts are identical to the ones used to generate the audio. This ensures cohesion between the host's dialogue and their voice. The process of mapping user selected voice parameters to an ElevenLabs voice ID is explained in the next section.

4.3 Audio Generation

The audio generation process for Creational and Premade podcasts includes both the podcast script and a voice ID to be submitted to the ElevenLabs API. The voice ID is obtained through a helper function, voiceMapper, within the Server module. This method includes all 90 host voices, which were manually selected from ElevenLabs' voice library. Where suitable voices were unavailable, they were manually created through ElevenLabs' descriptive voice creation.

Due to ElevenLabs' limitation in handling single-speaker text-to-speech requests, when students select a two-host podcast, multiple requests must be submitted to generate the podcast audio. This process involves splitting and submitting singular host lines and later merging the collection of audio clips together. Alternatively, if a single-speaker podcast is selected, only one request is submitted.

4.4 GAIDE

4.4.1 Manual

The GAIDE Manual workflow is implemented across four screens, designed to balance completeness with usability. Each screen aligns with a moderation step while keeping the workload

manageable for educators. To this end, Appendix F presents the set-up screen, Appendix G details the structure macro-refinement page, Appendix H details the script micro-refinement screen and lastly, Appendix I illustrates the final consolidation and upload screen. Across all refinement screens, the textbox presents the generated content to which the educator can edit.

4.4.2 Automatic

GAIDE Automatic features a single screen, presented in Appendix J, which requires the students to select their podcast parameters. The automatic moderation implementation in the form of its algorithm is outlined in Appendix K.

4.5 Compilation And Delivery

The compilation process involves merging the introductory theme song to the beginning of the audio file when the student has selected this option. The delivery of the audio file from the server to Anvil's client proceeds with uploading the file to a Google Cloud Storage bucket, the returned URL link is then passed along to Anvil, at this point, the user will simply click view and listen to the podcast. This optimization reduces the audio transmission from 40 seconds with Anvil Uplink to just 10 seconds on a two-minute podcast.

4.6 Optimization

On top of the aforementioned optimization in audio transmission, the sequential series of multi-speaker requests submitted to the ElevenLabs API are parallelized using multi-threading to further improve system responsiveness.

With these optimizations in place, Table 5 below presents the performance results for the total processing time, internal system processing time and internal processing share.

Table 5: Podcast System Performance.

Podcast Duration (minutes)	Total Processing Time (seconds)	System Processing Time (seconds)	Internal Processing Share
2	56.22	20.34	36.18%
4	75.80	27.50	36.27%
8	104.34	41.02	39.31%
10	141.36	63.16	44.68%

While the usability requirement in the form of system responsiveness is largely fulfilled, at higher podcast durations, the performance drops significantly. The lack of scalability is largely due to the transmission of increasingly large audio files. In addition, the system processing time and subsequently responsiveness depends on the user's internet bandwidth.

4.7 Other Features

In an effort to further boost student engagement, an alternative to manually selecting host settings is enabled by students creating their own AI host through descriptive text. This interactive feature allows students to create a truly immersive and personalized podcast. Real-time AI voice creation is enabled through ElevenLabs' descriptive voice generation API. An educator-supporting feature in the form of AI voice uploads is also included. This feature allows educators to upload new voices for use in Premade Podcasts, enabling the creation of new diverse content as ElevenLabs' voice library expands.

4.8 User Interfaces

The design choices made during the implementation of the system's user interfaces were directly informed by the usability requirement, which specified that at least nine out of Jakob Nielsen's ten heuristics [6] are followed. This subsection discusses how these principles were embedded in the core interfaces.

Nielsen's first and ninth heuristics state that the system's status should always be visible and that users should be supported in recovering from errors. This was enforced through information boxes that updated depending on system state and displayed fixes for errors. The information box was centrally located across all screens (Appendices F, I, and J). During text and audio generation, a loading circle also indicated system processing.

The second principle, matching the system to the real world, was supported through affordances and logical workflows, with icons representing actions (Appendices F–J). The third heuristic, user control and freedom, was enforced through back buttons and persistent navigation bars. Nielsen's fourth heuristic, consistency, was upheld by applying the same button functions across screens and following standard placement conventions (Appendices G–I).

Error prevention was achieved through dropdowns, radio buttons, and sliders that restricted inputs and reduced mistakes. Sequential screens during the Premade process enforced recognition rather than recall by only presenting stage-relevant actions.

The seventh heuristic, flexibility and efficiency of use, was not fully supported, as both the Creational and Premade processes required explicit interactions; shortcuts were not incorporated. The eighth heuristic, aesthetic and minimalistic design, was reinforced through clean, focused interfaces. Lastly, system tutorials embedded in information boxes fulfilled the final principle by providing help and documentation (Appendices G, I, and J).

To this end, the system met the usability requirement, ensuring alignment with Nielsen's heuristics [6] and providing both students and educators with accessible and supportive user interfaces.

5 USER TESTING

5.1 Materials

To evaluate the research question, "To what extent do first-year science students perceive the AI-generated podcast as more engaging than traditional student guides?", the student engagement survey was employed. This survey was based on the Situational Interest Survey for Multimedia (SIS-M) [7] described in the CNN related works but adapted to align more closely with the objectives of this study. While items one to four measured students' engagement with the podcast, items five to seven evaluated the alignment of the podcast with the original student guide as well as the content itself. The adapted version is presented in Appendix L.

In addition, to investigate "What AI agent personalities and demographics do first-year science students prefer", the student preference survey was used. This survey included participant demographic items as well as the podcast parameters selected by students during the experiment. The survey is presented in Appendix M.

In terms of student participation, invitations were sent via email to first-year science student groups, with the process beginning only after ethical clearance had been obtained.

5.2 Methodology

The experiment began with participants completing the student preference survey. Following this, students read the *So, How Does Your Brain Work?*[18] student guide for five minutes, followed by filling out the student engagement survey. Students were then introduced to the podcast system, where they selected the formality and host settings for their podcast. The student guide and length parameters were fixed to the same student guide that was read and a duration of two minutes. This ensured that students could fairly compare the original student guide to the new system. The two-minute podcast length was also chosen to minimize participant fatigue, as each student went on to complete two additional, similar experiments. After the podcast was generated, participants listened to it in full and then completed the student engagement survey a second time. The placement of the student preference survey at the start of the experiment helped reduce bias, as after students read the student guide for five minutes, their podcast settings selections were less likely to be influenced by their demographic responses.

6 RESULTS

Following the testing methodology presented, sixteen first-year science students were recruited for participation. Appendix N presents the Student Engagement Results (S1–S7, N=16), detailing individual and average scores for the Original Student Guide (O) and the Podcast System Iteration 1 (I1). Appendix O summarises these results with the mean, standard deviation (SD), standard error of the mean (SEM), and paired *t*-test outcomes used to assess statistical significance. Together, Appendices N and O capture the responses to all seven survey statements (Appendix L). Lastly, Appendix P provides the Student Preference Results, including participant demographics and podcast parameter selections.

6.1 Student Engagement

The student engagement bar chart**,** Figure 6 below compares the average score for each statement (S1-S7) under the Original Student Guide (O) and the Podcast System Iteration 1 (I1), Student Engagement Results (Appendix N).

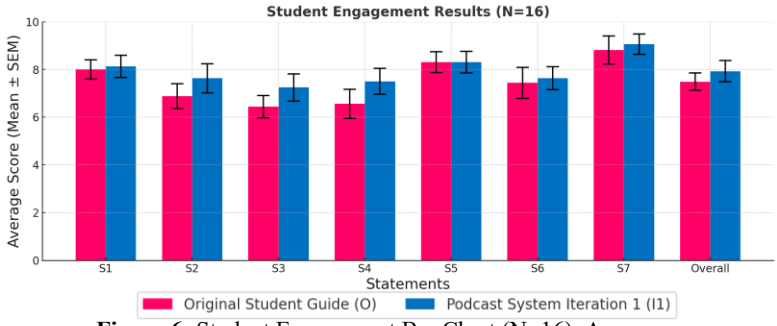


Figure 6: Student Engagement Bar Chart (N=16). Average scores (Mean ± SEM) for each survey statement (S1–S7).

Across all statements, the first iteration of the podcast system consistently achieved higher average scores compared to the original student guide, with the largest improvements observed in S2–S4. Statement 5 showed identical average scores across both guides, while S1 saw only a marginal increase. The overall average increased from **7.49 ± 0.36 (O)** to **7.93 ± 0.45 (I1)**.

While the overall average showed a modest 0.44-point increase, a further breakdown indicates a larger 0.66-point increase across statements related to student engagement (S1–S4). The smaller 0.14-point increase across statements S5–S7 shows that the podcast remained closely aligned with the original student guide content.

However, none of the individual statements reached statistical significance (all $p > 0.26$), and the overall comparison similarly did not reach significance ($t = -0.71$, $p = 0.49$).

6.2 Student Preference

Figure 7 below summarizes student preferences for podcast formality, host age, accent, gender, tone, and style.

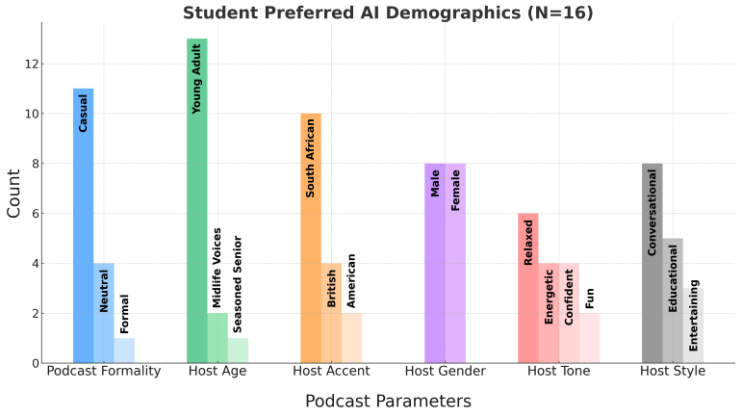


Figure 7: Student Preferred AI Demographics (N=16)

Figure 7 shows that most students preferred a casual formality (11/16) and a young adult host (13/16) with a South African accent (10/16). Gender preferences were evenly split (8/16 each). A relaxed tone (6/16) and conversational style (8/16) were most common, with smaller groups selecting energetic, confident, or entertaining podcast options.

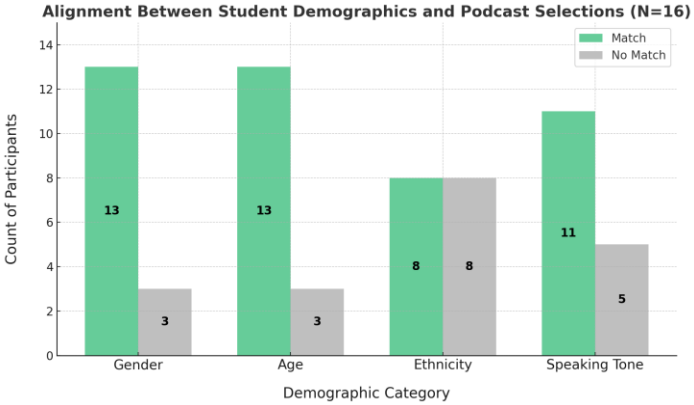


Figure 8: Alignment between student demographics and podcast selections (N=16).

In addition, Figure 8 above presents the relationship between student demographics and their chosen podcast parameters. It can be observed that most students preferred AI host age and gender settings that aligned with their own (13/16). On the other hand, students showed no clear preference for host ethnicity and only a smaller group selected a speaking tone that matched their own (11/16).

7 DISCUSSION

This paper aimed to investigate the two research questions: whether first-year science students perceive AI-generated podcasts as more engaging than traditional student guides, and what host demographics and personalities they prefer. While the overall improvement of 7.49 ± 0.36 (O) to 7.93 ± 0.45 (I1) indicates that the podcast system does increase student engagement, a further breakdown of these results reveals additional insights.

The student engagement results can be broken down into two categories: whether students perceived the podcast system as more engaging than the original student guide as indicated by statements one to four. In addition, statements five to seven represent how closely aligned the AI-generated podcasts were to the original student guide as well as the content itself. While neither individual nor overall scores achieved statistical significance due to the small sample size, the consistent positive trend across statements suggests that further testing with a larger student group may provide more conclusive results.

The larger 0.66-point increase across statements one to four affirms the paper's hypothesis that students will report higher levels of engagement when using the proposed student guide compared to the existing PDF-based student guide. This can be attributed to three factors. The first and largest being the personalization aspect of the podcast system. By allowing students to directly control the generation of their podcast, the system enhances this learning process by enabling creativity and interactivity. Furthermore, several participants commented that the ability to control the length of the podcast enhances its accessibility as shorter podcasts reduced cognitive load and longer podcasts provided detailed information. This is further complemented by the second factor, access to a diverse range of host voices. By allowing students to create their own hosts, the generated podcasts were more relatable and created a sense of ownership for their learning. Lastly, the entire process is facilitated by a minimalistic and highly usable user interface which reduced cognitive effort and allowed students to focus on the learning content rather than the system itself. These results largely align with related works as both CNN and Personalized Programming Exercises both reported higher student engagement through the use of personalization and generative AI.

However, the podcast system had several limitations. The first limitation was that several host voices were unemotive, monotone or misaligned with their description, particularly the South African voice group. This can be observed in the results (Appendix N and P) where students who selected British voices reported higher engagement scores compared to those who chose a South African voice. This can be attributed to British voices having a greater representation in ElevenLabs' training data. On the other hand, South African accents were less well modeled, leading to reduced engagement as students noted mispronunciations of traditional phrases such as "hakuna-matata". This was also observed in the CNN study, where students felt underrepresented when their culture wasn't adequately reflected in the system's output [7]. Participants also expressed concern about falling behind during the podcast, whereas the traditional student guide allowed them to learn at their own pace. Lastly, several students noted the absence of visual elements which reduced their engagement and learning effectiveness as they themselves were visual learners.

Statements five to seven showed a marginal increase of 0.14 points. This indicates that the podcast system generated scripts that were aligned with the original student guide and occasionally provided more comprehensive explanations. This can be attributed to the GAIDE Automatic framework as it ensured relevancy and alignment through iterative refinement. This is facilitated by the underlying generative-AI pipeline which separated the generation of the podcast script and audio. In addition, the increase is a result of the script generation process, which incorporated analogies to better explain concepts. However, students commented on the low information density, noting that the two-minute podcast constraint during testing reduced the amount of content delivered, as a portion of the time was taken up by the introduction and outro.

The student preference results also affirm the paper's hypothesis that students will prefer AI host demographics and personalities that align with their own characteristics. This aligns with a study by Zhang et al. (2025) [19], which found that users favored AI voices that align with their own gender identity, revealing how demographic alignment can support trust and engagement.

To this end, the final iteration will address these limitations with two refinements. The first of which will replace ElevenLabs `eleven_multilingual_v2` with the recently released `eleven_v3`. This will address the host's monotone delivery during the podcast. Secondly, the minimum podcast length will be increased from one to three minutes to enable more effective delivery of information.

8 CONCLUSION

Through an iterative design process, this paper presented an AI-generated podcast system as an alternative to traditional student guides. The system achieved its aim of demonstrating that personalization, diverse host options, and a usable interface can support student engagement beyond that of traditional student guides. This was reinforced by the GAIDE moderation framework, which ensured that generated podcasts remained aligned with the original content. In addition, the paper found that students prefer AI-host demographics that match their own, highlighting the importance of relatability and trust within learning tools. However, limitations such as the short podcast duration, uneven voice quality, and the absence of visual support reduced its effectiveness. As such, future extensions will include experiments with larger sample sizes to validate the paper's findings and refinements to accent modelling, particularly for underrepresented and dataset-limited voice groups. Furthermore, simple improvements such as dialogue-speed controls and host-voice previews will enable students to learn at their own pace and pick suitable voices before generating a full podcast. Beyond usability, systematic testing of GAIDE Automatic's iteration cap, scoring threshold and effectiveness will establish its role in AI-enabled learning environments. This paper also presented two podcast generation modes, Creational and Premade, highlighting efficiency versus full validation tradeoffs. Subsequently, future work could identify a middle ground to balance these tradeoffs. While integrating visual elements would reduce the podcast's accessibility advantage, minimal graphical and textual elements may enhance student learning and engagement. To this end, the podcast system shows promise for supporting active learning, but it may currently be better suited as a reinforcement medium, where students can consolidate their understanding after they have read the student guide PDFs.

9 REFERENCES

[1] Mohammed Kajee, Dale Taylor, Olerato Mogomotsi, Tsatsi Mnisi, Denzel Mtoko, and Akha Tutu. 2024. Culture Shock at UCT. University of Cape Town. DOI:https://doi.org/10.25375/uct.25551108.v2

[2] Mohammed Kajee, Dale Taylor, and Nomsa Ledwaba. 2025. *Orientation.* University of Cape Town. 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. DOI:https://doi.org/10.1016/j.sbspro.2013.12.625

[4] Glenda Barlow-Jones. 2019. The Struggles Experienced by First-Year Computer Programming Students at a University in South Africa. In *Proceedings of the 2nd International Conference on Research in Education*. DOI:https://doi.org/10.33422/2nd.icre.2019.12.952

[5] Eliza Lily Goode, Johanna E. Nieuwoudt, and Trish Roche. 2022. Does online engagement matter? The impact of interactive learning modules and synchronous class attendance on student achievement in an immersive delivery model. *Australasian Journal of Educational Technology* 38, 4 (2022), 76–94. DOI:https://doi.org/10.14742/ajet.7929

[6] Jakob Nielsen. 2006. Ten Usability Heuristics. Nielsen Norman Group. Retrieved March 21, 2025 from https://www.nngroup.com/articles/ten-usability-heuristics/

[7] Tyler Bland. 2025. Enhancing Medical Student Engagement Through Cinematic Clinical Narratives: Multimodal Generative AI–Based Mixed Methods Study. *JMIR Medical Education* 11 (2025), e63865. DOI:https://doi.org/10.2196/63865

[8] Leonardo.ai. 2025. AI image generator – create art, images & video. Retrieved March 21, 2025 from https://leonardo.ai.

[9] ElevenLabs. 2025. AI voice generator – text to speech & voice cloning. Retrieved March 10, 2025 from https://elevenlabs.io.

[10] Tonia A. Dousay. 2016. Effects of redundancy and modality on the situational interest of adult learners in multimedia learning. *Educational Technology Research and Development* 64, 6 (2016), 1251–1271. DOI: https://doi.org/10.1007/s11423-016-9456-3.

[11] Tonia A. Dousay and Nicholas P. Trujillo. 2019. An examination of gender and situational interest in multimedia learning environments. *British Journal of Educational Technology* 50, 2 (2019), 876–887. DOI: https://doi.org/10.1111/bjet.12610.

[12] Tyler Bland, Min Guo, and Tonia A. Dousay. 2024. Multimedia design for learner interest and achievement: A visual guide to pharmacology. *BMC Medical Education* 24, 1 (2024), 113. DOI: https://doi.org/10.1186/s12909-024-05077-y

[13] Evanfiya Logacheva, Arto Hellas, James Prather, Sami Sarsa, and Juho Leinonen. 2024. Evaluating Contextually Personalized Programming Exercises Created with Generative AI. In *Proceedings of the 2024 ACM Conference on International Computing Education Research (ICER '24)*, Vol. 1. ACM, New York, NY, 95–113. DOI:https://doi.org/10.1145/3632620.3671103

[14] Mark Guzdial. 2015. *Learner-Centered Design of Computing Education: Research on Computing for Everyone.* Morgan & Claypool Publishers, San Rafael, CA. DOI: https://doi.org/10.2200/S00684ED1V01Y201511HCI033

[15] Riikka Leinonen, Ville Isomöttönen, and Matti Tedre. 2017. Reducing mathematics-related stereotypes through contextualized programming. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '17).* ACM, New York, NY, USA, 142–147. DOI: https://doi.org/10.1145/3059009.3059039

[16] Juan Carlos del Carpio Gutierrez, William Gacitúa, and Jorge Pérez. 2023. Evaluating the Quality of Programming Problems Generated by Large Language Models Across Multiple Languages. arXiv preprint arXiv:2304.11045.

[17] 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)*. IEEE, 1–9.

[18] D. Taylor and C. Mahony. 2024. *So, How Does Your Brain Work?* University of Cape Town, Educational resource. DOI:https://doi.org/10.25375/uct.25592541.v2

[19] Y. Zhang, X. Li, and H. Chen. 2025. Artificial Intelligence Voice Gender, Gender Role Congruity and Voice Preference. *Computers in Human Behavior* 154 (2025), 108223. DOI:https://doi.org/10.1016/j.chb.2024.108223.

APPENDIX

Appendix A: SIS-M Survey Type and Items

SIS[a] type	Survey item
SI-triggered	The multimedia presentation was interesting.
SI-triggered	The multimedia presentation grabbed my attention.
SI-triggered	The multimedia presentation was often entertaining.
SI-triggered	The multimedia presentation was so exciting, it was easy to pay attention.
SI-maintained-feeling	What I learned in the multimedia presentation is fascinating to me.
SI-maintained-feeling	I am excited about what I learned in the multimedia presentation.
SI-maintained-feeling	I like what I learned in the multimedia presentation.
SI-maintained-feeling	I found the information in the multimedia presentation interesting.
SI-maintained-value	What I studied in the multimedia presentation is useful for me to know.
SI-maintained-value	The things I studied in the multimedia presentation are important to me.
SI-maintained-value	What I learned in the multimedia presentation can be applied to my job.
SI-maintained-value	I learned valuable things in the multimedia presentation.

[a]SIS: Situational Interest Survey.

Appendix B: Preliminary Evaluations of Text-to-speech Technologies

Rated from 1 to 10 with 1 indicating lowest and 10 indicating highest.

Platform	**Expressivenss**	**Stability**	**Range of Voices**	**Ease of Implementation**	**Affordability**	**Multi-Speaker Capability**
PlayHT	9.5	8.0	7.5	10.0	8.0	Yes
ElevenLabs	7.5	9.0	8.0	9.0	8.0	No
Microsoft Azure Cognitive TTS	5.5	9.0	9.0	7.0	8.0	No
Google Cloud TTS	5.5	9.0	9.0	6.0	8.0	No
Amazon Polly	6.0	8.5	9.0	5.0	9.0	No

Appendix C: Creational Podcasts Structured Prompt

```
You are an expert podcast scriptwriter. Produce a student-focused educational episode script
that is highly conversational and exactly follows these specifications:

1. Topic & Length
   • Target length: approximately {length_minutes} minutes.

2. General Tone & Style
   • Language: {language}
   • Formality: {formality}
   • {"Use occasional humor and rhetorical questions."
        if formality.lower() in ("casual","neutral")
        else "No humor; keep tone strictly {formality.lower()}."}

3. Host Details
   We have {num_hosts} host(s):
   • Host 1: Age: {host1_age}, Accent: {host1_accent},
             Gender: {'male' if host1_gender else 'female'},
             Tone: {host1_tone}, Style: {host1_style}.
   {if num_hosts == 2:
        Host 2: Age: {host2_age}, Accent: {host2_accent},
                Gender: {'male' if host2_gender else 'female'},
                Tone: {host2_tone}, Style: {host2_style}.}

4. Structure
   a. Hook
   b. Introduction
   c. Key Points
   d. Conclusion

5. Formatting
   - Always label each line “Host1:” or “Host2:”.
   - No extra headers, sound cues, or markup.
   - Use short sentences and rhetorical questions where allowed.

Podcast Episode Name:
   {student_guide_title}

Source Content (use this for facts & examples):
   {student_guide}

Begin script now, strictly following the above.
```

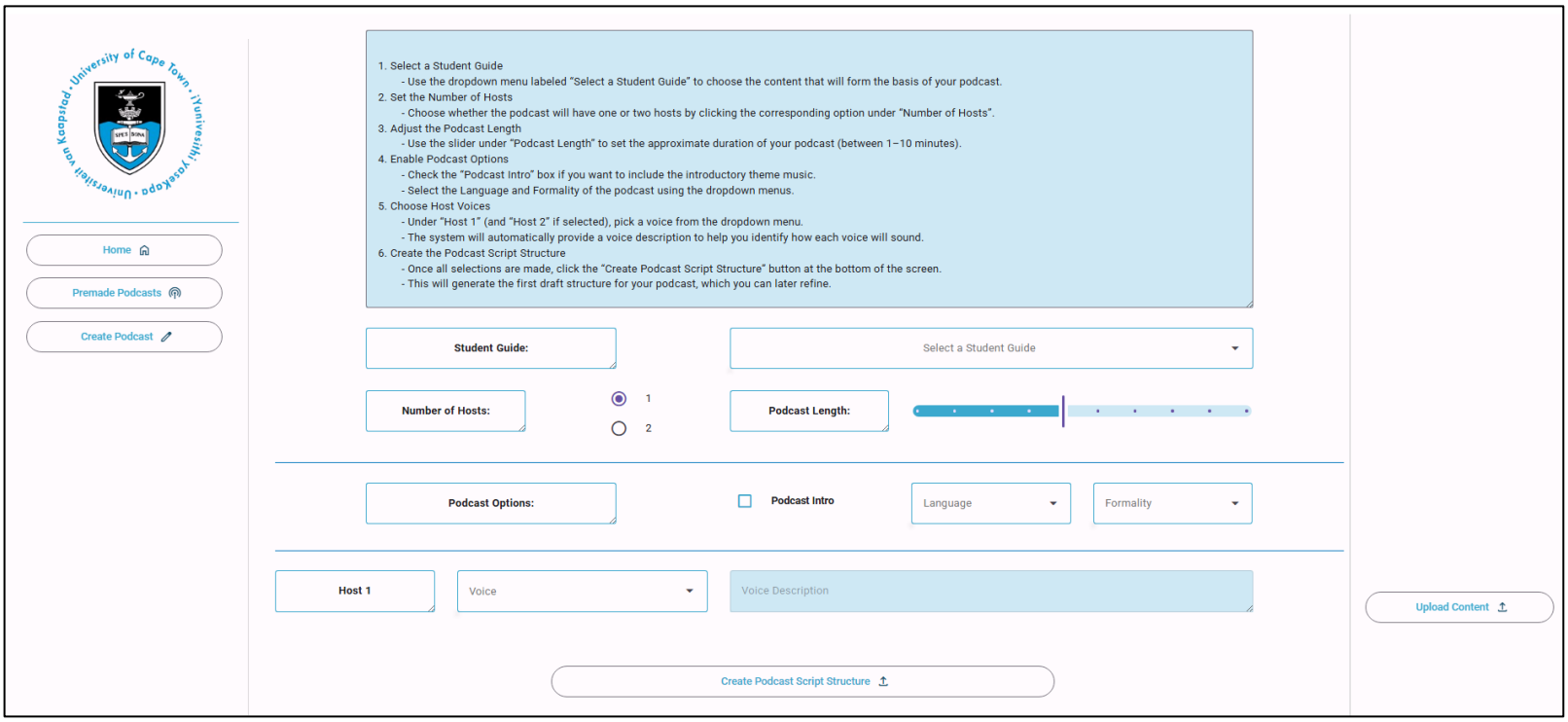
Appendix D: Premade Podcasts Structured Prompt – Structure

```
You are a podcast showrunner. Create a high-level structure for a student-focused educational podcast episode.
Do NOT write dialogue. The output must be plain text with section names, objectives, and talking point summaries.

### Episode Info
- Topic/Title: {student_guide_title}
- Total Duration: ~{length_minutes} minutes (for pacing guidance only; do not display in output)
- Language: {language}
- Formality: {formality}
- Hosts ({num_hosts} total): {", ".join(hosts_desc)}
- You may include an Intro if it helps the flow.

### Source Material
Use this ONLY to inspire the topics, objectives, and flow:
{student_guide}

### Output Format (Plain Text)
Episode Title: <refined title if needed>

Formality: <formality> | Language: <language>
Hosts: <host descriptions>

Section 1: <Name>
Objective: <one line>
Summary: <1–2 sentences>

Section 2: <Name>
Objective: <one line>
Summary: <1–2 sentences>

(...continue until the episode feels complete...)

Conclusion
Summary: <1–2 sentences>
Takeaway: <final message or call to action>

Rules:
- Keep bullets short and scannable.
- No dialogue or filler.
- The number of sections is up to you.
- Ensure that the relevancy and focus of the structure aligns with the student guide.
```

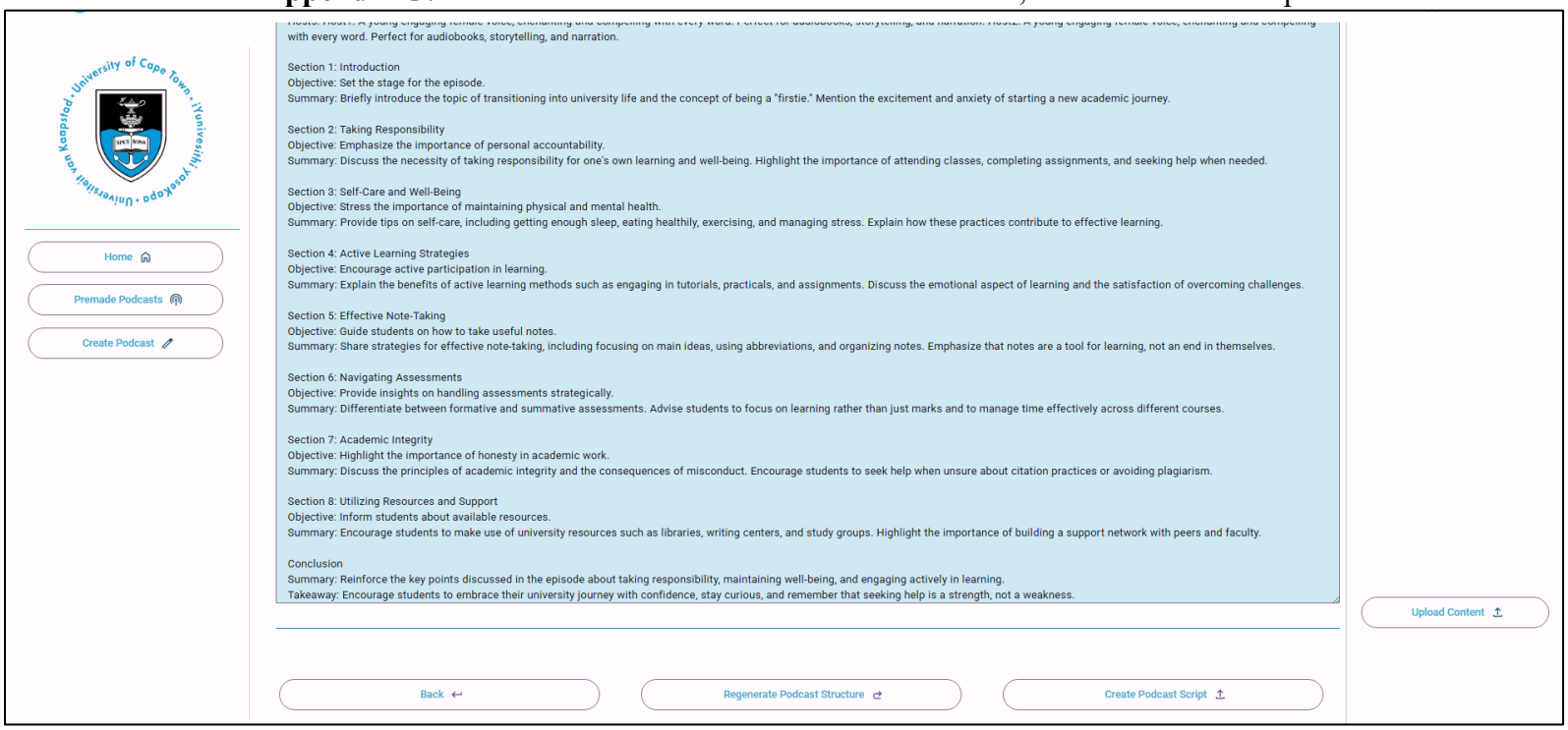
Appendix E: Premade Podcasts Structured Prompt – Script

```
You are an expert podcast scriptwriter. Turn the following podcast structure into a
fully written, student-focused, educational dialogue script.

### Topic & Length
- Topic/Title: {student_guide_title}
- Target length: approximately {length_minutes} minutes (guidance only; do not output timestamps).

### General Tone & Style
- Language: {language}
- Formality: {formality}
- {humor_clause}

### Host Details
{hosts_block}

### Script Structure to Follow
{script_structure}

### Important
- Ensure that the relevancy and focus of the structure aligns with the student guide.

### Output Rules
- Output only dialogue lines starting with "Host1:" or "Host2:" exactly.
- No headings, section labels, timestamps, stage directions, or sound effects.
- Keep sentences short and conversational.
- Base all content on the provided structure and this source material:
{student_guide}
{host_rules}

Begin the dialogue now.
```

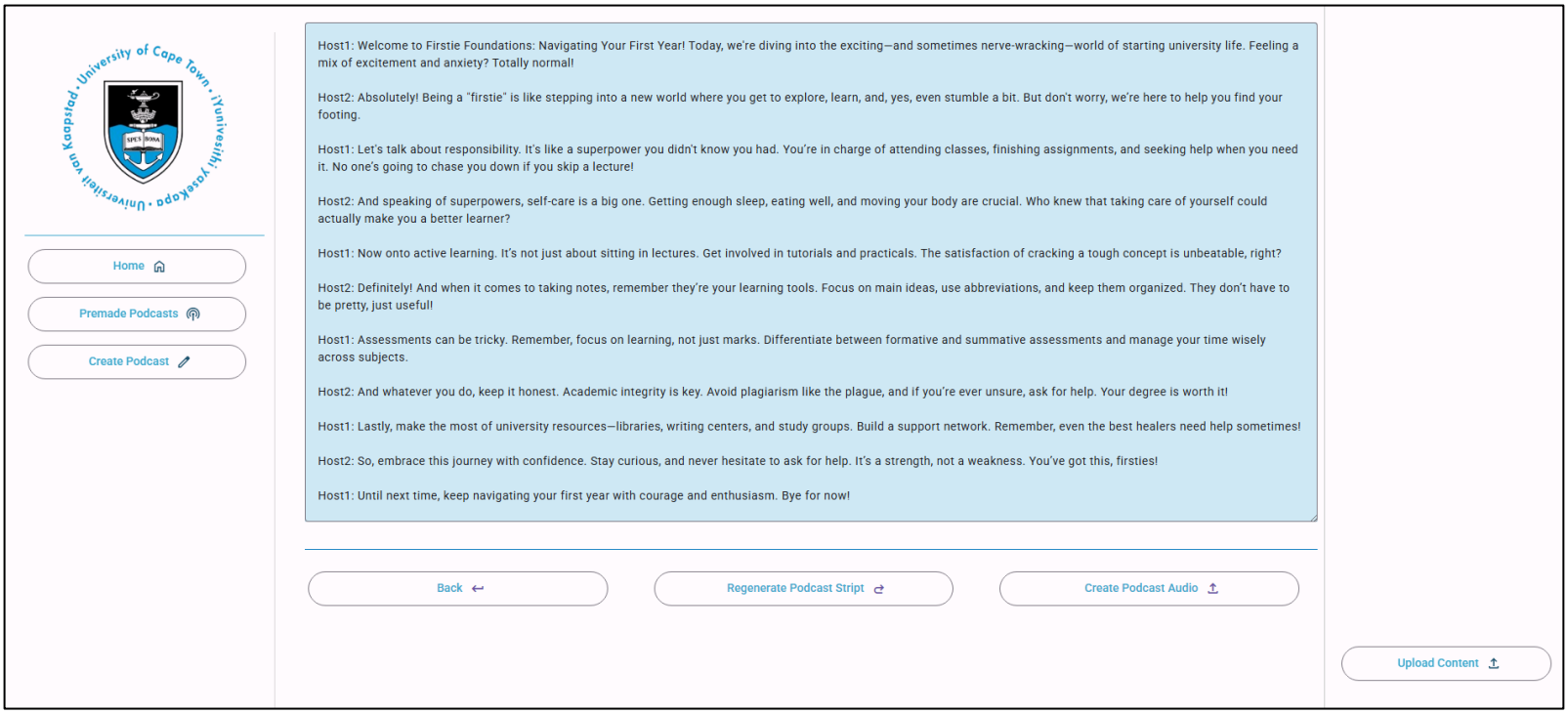
Appendix F: Premade Podcast - Setup Screen



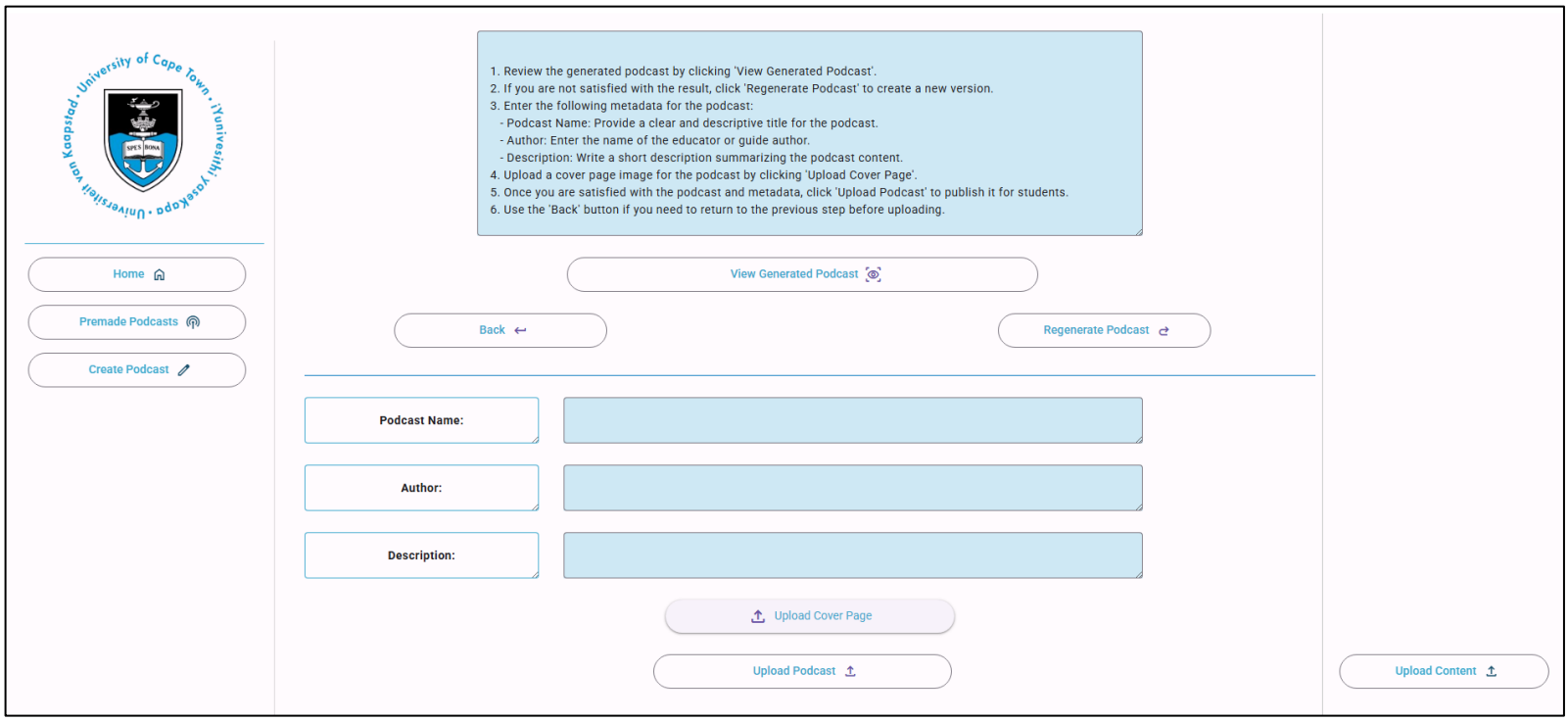
Appendix G: Premade Podcast - Macro-Refinement Screen, with structure example.



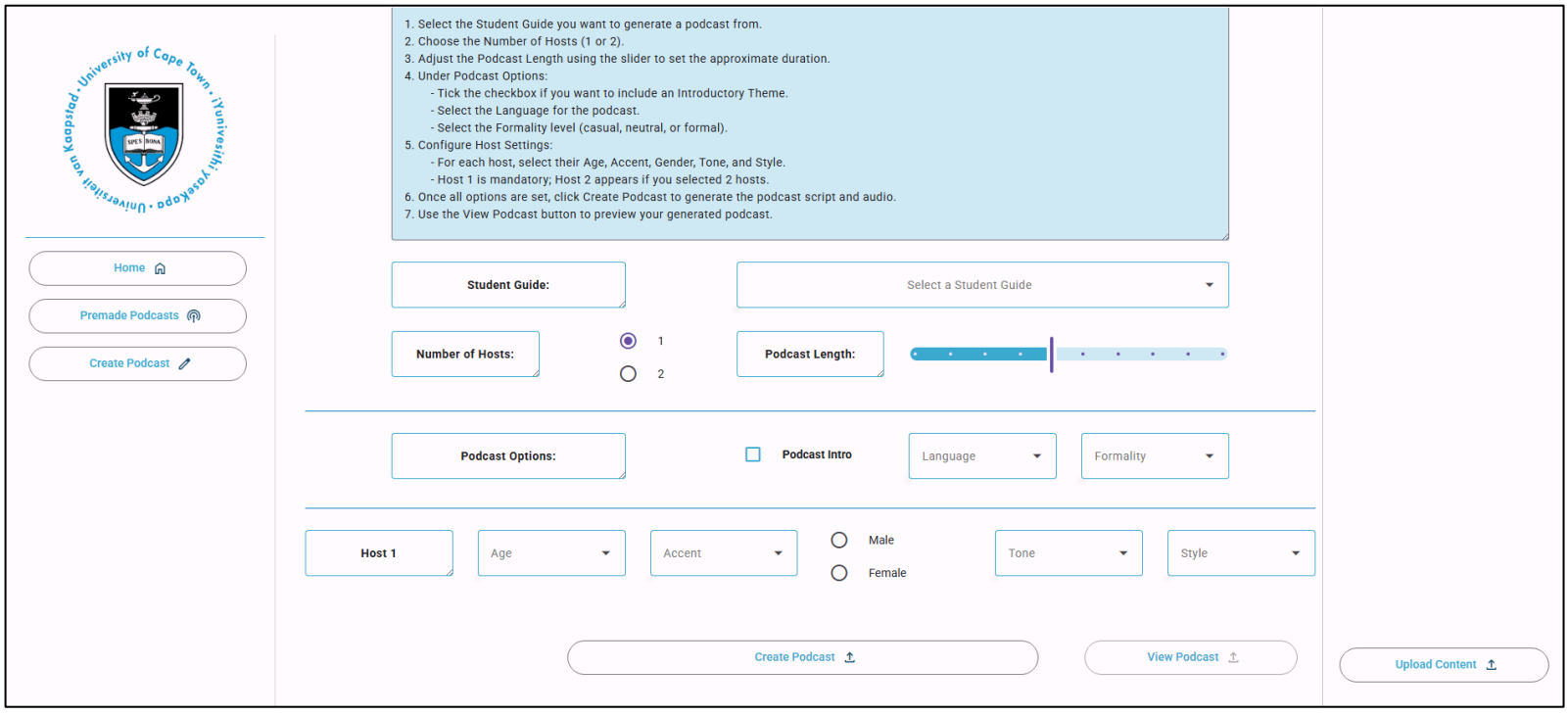
Appendix H: Premade Podcast – Micro-Refinement Screen, with script example.



Appendix I: Premade Podcast – Consolidation and Upload Screen



Appendix J: Creational Podcast – Setup Screen



Appendix K: GAIDE Automatic Algorithm

```
METHOD GAIDE_Automatic(script, guide_content, target_score, max_rounds):

    rounds_used = 0
    best_script = script
    best_score = -1
    best_feedback = ""

    WHILE rounds_used < max_rounds DO
        rounds_used = rounds_used + 1

        // Step 1: Ask ChatGPT to evaluate the script
        evaluation_prompt = "Evaluate the following script against the guide content.
                             Score 0–100 based on accuracy, alignment, clarity, and tone.
                             Return both a score and short feedback.
                             SCRIPT: " + script + "
                             GUIDE CONTENT: " + guide_content
        evaluation = ChatGPT(evaluation_prompt)

        score = PARSE_SCORE(evaluation)
        feedback = PARSE_FEEDBACK(evaluation)

        IF score IS NULL THEN
            score = 0
        ENDIF

        IF score > best_score THEN
            best_score = score
            best_script = script
            best_feedback = feedback
        ENDIF

        // Step 2: If score meets threshold or max rounds reached, stop
        IF score >= target_score OR rounds_used = max_rounds THEN
            BREAK
        ENDIF

        // Step 3: Refine script using feedback and guide content
        refine_prompt = "Refine the following script using this feedback: " + feedback + "
                         Ensure the revision remains aligned with the guide content,
                         improves quality, and does not introduce new errors.
                         SCRIPT: " + script + "
                         GUIDE CONTENT: " + guide_content
        script = ChatGPT(refine_prompt)

    ENDWHILE

    RETURN (best_script, best_score, best_feedback)
END METHOD
```

Appendix L: Student Engagement Survey

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.										

Appendix M: Student Preference Survey

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 capture student information in evaluating what AI agent demographics - age, gender, ethnicity and language - first year science students prefer.

Name: **Student Number:**

No.	Statement					
1	Students age.					
2	Students' ethnicity.					
3	Students gender.					
4	Students' primary language.					
5	Students preferred tone during conversations.	Fun	Relaxed	Professional	Confident	Energetic

Below are the AI agent demographics chosen by the students during the experiments, it will be filled out by the researcher.

Researcher: **Student Number:**

No.	Podcast Parameter	
1	Student Guide content.	
2	Number of hosts.	
3	Podcast length.	
4	Podcast language.	
5	Podcast formality.	
6	Host age.	
7	Host accent.	
8	Host gender.	
9	Host tone.	
10	Host style.	

Appendix N: Student Engagement Results (S1-S7, N=16)

S1–S7 correspond to the seven statements in the Student Engagement Survey presented earlier. O indicates responses for the Original Student Guide, while I1 indicates responses for the first system iteration.

	S1		S2		S3		S4		S5		S6		S7	
Participants	**O**	**I1**	**O**	**I1**	**O**	**I1**	**O**	**I1**	**O**	**I1**	**O**	**I1**	**O**	**I1**
P1	8	6	8	8	6	6	6	8	8	10	8	8	10	10
P2	10	10	10	10	8	10	10	8	10	10	10	10	10	10
P3	8	10	5	10	10	8	5	10	3	10	1	8	1	10
P4	7	5	7	6	5	6	5	6	9	7	10	8	10	9
P5	10	8	10	9	9	8	10	7	10	8	10	7	10	10
P6	9	6	4	4	6	6	7	5	10	8	7	7	8	7
P7	8	6	7	2	6	3	7	2	8	4	7	4	9	7
P8	8	9	7	10	5	8	8	7	7	8	4	8	10	10
P9	7	5	9	4	7	2	6	6	10	5	9	3	10	4
P10	8	9	5	8	4	9	2	7	8	8	6	8	9	9
P11	8	8	6	7	6	8	5	7	8	8	7	8	10	9
P12	8	9	7	8	9	9	7	8	10	10	9	10	10	10
P13	10	10	8	8	6	6	5	9	9	10	9	7	10	10
P14	8	9	5	9	7	8	10	10	7	8	4	7	6	10
P15	8	10	9	9	6	9	9	10	8	9	10	9	10	10
P16	3	10	3	10	3	10	3	10	8	10	8	10	8	10
Average(μ)	8.00	8.12	6.88	7.62	6.44	7.25	6.56	7.50	8.31	8.31	7.44	7.62	8.81	9.06

Appendix O: Statistical Summary of Student Engagement Survey (S1–S7, N=16)

Reported values include mean, standard deviation (SD), standard error of the mean (SEM), and paired t-test significance for Original (O) vs Iteration 1 (I1).

	Mean		SD		SEM		Paired t-test (N=16)	
Statement	**O**	**I1**	**O**	**I1**	**O**	**I1**	***t*-stat**	***p*-value**
S1	8.00	8.13	1.63	1.89	0.41	0.47	-0.20	0.841
S2	6.88	7.63	2.09	2.45	0.52	0.61	-0.94	0.362
S3	6.44	7.25	1.86	2.29	0.47	0.57	-1.10	0.288
S4	6.56	7.50	2.42	2.16	0.61	0.54	-1.17	0.262
S5	8.31	8.31	1.78	1.82	0.44	0.45	0.00	1.000
S6	7.44	7.63	2.61	1.93	0.65	0.48	-0.24	0.814
S7	8.81	9.06	2.37	1.69	0.59	0.42	-0.32	0.751
Overall Results	7.49	7.93	±1.42	±1.82	±0.36	±0.45	-0.71	0.49

Appendix P: Student Preference Results (N=16)

Speaking tone refers to the student's preferred tone during conversations.

	Student Demographics				Podcast Parameters Selected					
Participants	**Age**	**Ethnicity**	**Gender**	**Speaking Tone**	**Podcast Formality**	**Host Age**	**Host Accent**	**Host Gender**	**Host Tone**	**Host Style**
P1	19	South African	Male	Energetic	Formal	Young Adult	South African	Male	Energetic	Educational
P2	18	South African	Male	Fun	Casual	Young Adult	South African	Male	Relaxed	Conversational
P3	18	South African	Female	Relaxed	Casual	Young Adult	South African	Male	Energetic	Conversational
P4	19	South African	Male	Relaxed	Casual	Young Adult	South African	Male	Relaxed	Conversational
P5	19	South African	Female	Energetic	Casual	Young Adult	South African	Female	Energetic	Educational
P6	18	South African	Female	Relaxed	Neutral	Young Adult	South African	Female	Relaxed	Conversational
P7	19	Indian	Female	Fun	Neutral	Midlife Voices	South African	Female	Confident	Conversational
P8	19	South African	Male	Confident	Casual	Young Adult	British	Male	Confident	Educational
P9	18	South African	Male	Professional	Casual	Young Adult	South African	Female	Confident	Educational
P10	19	South African	Female	Fun	Casual	Young Adult	British	Female	Fun	Conversational
P11	18	Indian	Male	Relaxed	Neutral	Young Adult	South African	Male	Relaxed	Conversational
P12	18	South African	Male	Confident	Casual	Midlife Voices	American	Male	Confident	Entertaining
P13	19	South African	Male	Relaxed	Casual	Young Adult	British	Female	Fun	Conversational
P14	19	South African	Female	Energetic	Casual	Young Adult	South African	Female	Energetic	Entertaining
P15	18	South African	Female	Relaxed	Neutral	Young Adult	British	Female	Relaxed	Entertaining
P16	19	South African	Male	Relaxed	Casual	Seasoned Senior	American	Male	Relaxed	Educational