IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY

BY

NG JING YING

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology
(Kampar Campus)

JAN 2024

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY

Academic Session: __JAN 2024___

I ___________________NG JING YING________________________

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature) (Supervisor's signature)

Address:

83, Jalan Pahlawan 3, Taman Ungku Tun Aminah, 81300 Skudai, Johor.

Prof Ts Dr Liew Soung Yue

Supervisor's name

Date: 25 April 2024 **Date**: 26/4/2024

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 25 April 2024

SUBMISSION OF FINAL YEAR PROJECT

It is hereby certified that **NG JING YING** (ID No: **21ACB01845**) has completed this final year project entitled "**IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY**" under the supervision of **Prof Ts Dr Liew Soung Yue** (Supervisor) from the **Department of Computer and Communication Technology**, **Faculty of Information and Communication Technology**.

I understand that University will upload softcopy of my final year project in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

(*NG JING YING*)

DECLARATION OF ORIGINALITY

I hereby declare that this report entitled "**IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : ______________________________

Name : Ng Jing Ying

Date : 25 April 2024

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my supervisor, Prof Ts Dr Liew Soung Yue who has given me this opportunity to engage in a Large Language Model-based project. It is my first significant step towards establishing a career in the Generative Artificial Intelligence field. I am truly thankful for his mentorship in the journey to develop a novel solution under his experienced guidance.

Apart from that, I would like to extend my deepest thanks to my parents and family for their unwavering love and continuous encouragement throughout my academic journey. Their support has driven me to successfully complete this project on time.

ABSTRACT

As Malaysia undergoes a significant transition towards a skills-based economy, there is an increasing demand for soft skills training courses as individuals seek to gain their job competencies.

An immersive soft skills training application involves delivering scenario-based simulation practices and personalized feedback. However, existing Virtual Reality (VR) training applications are still struggling to balance cost-effectiveness, cognitive realism and comprehensive evaluation. This is because most existing applications rely on a decision-tree approach, where the storyline is constrained by preset branching choices. This not only requires a lot of human input to complete the storyline, but the overall experience still lacks cognitive realism. Besides, most existing applications only rely on quantitative metrics for evaluation, which fall short of providing comprehensive feedback in terms of soft skills training.

In this project, the main objective is to develop an immersive soft skills training application using Large Language Models (LLMs) and VR. In short, we have demonstrated the capability of LLMs in generating human-like behaviours. Besides, the combination of quantitative and qualitative data has improved the comprehensiveness of the evaluation process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
DSSC	Department of Soft Skills Competency
GPT	Generative Pretrained Transformer
IT	Information Technology
LLM	Large Language Model
NLP	Natural Language Processing
RAG	Retrieval Augmented Generation
VR	Virtual Reality

Chapter 1 Introduction

1.1 Problem Statement and Motivation

As Malaysia undergoes a significant transition towards a skills-based economy, Koo [1] emphasised that maintaining competitiveness requires proficiency in both soft and hard skills. This increases the demand for soft skills training modules. To offer soft skills courses that can concurrently accommodate many trainees, the online mode can be considered, however, the effectiveness of online-based modules is limited due to the potential distractions and the lack of in-person interaction among trainees [2]. To address these shortcomings, many educational technology providers in Western countries, such as Virtual Speech [3], Mursion [4] and STRIVR [5] have started incorporating Virtual Reality (VR) and Artificial Intelligence (AI) to virtually coach trainees in a simulated environment. A study [6] conducted in the US showed that VR training tends to be more time-saving and cost-effective, especially when implemented at scale.

An immersive soft skills training application involves delivering scenario-based simulation practices and personalized feedback. However, existing Virtual Reality (VR) training applications are still struggling to balance cost-effectiveness, cognitive realism and comprehensive evaluation.

Problem Statement 1: Most existing applications rely on a decision-tree approach, where the storyline is constrained by preset branching choices.

Most creation of VR content in existing applications requires a lot of human input to design and implement every possible outcome in response to trainees' actions. As a result, scene creation becomes a time-consuming process. Besides, such an approach reduces the cognitive realism of the scenario as the trainees may simply follow the provided options to progress in the simulation practice instead of reacting spontaneously.

Problem Statement 2: Most existing applications only rely on quantitative metrics for evaluation.

There are many applications like [3] and [7] that claim to improve communication skills by providing simulation practices like public speaking and virtual interviews. Unfortunately, most of the existing applications only focus on quantitative analysis, for instance, metrics such as speech rate, speech tone and speech loudness are the commonly used data to evaluate the learning outcomes. This is due to the limitation of the underlying models, as traditional rule-based methods or machine learning methods have limited context-understanding capability. As a result, such an approach fails to assess different soft skills values like assertive communication skills.

Motivation

While successful examples exist, there is still room for improvement in the existing VR training systems. Recently, a trending Large Language Model (LLM), ChatGPT has gained the attention of researchers and developers to boost any Natural Language Processing (NLP) driven use cases.

VR soft skills training is still a new area of research, especially when Natural Language Processing (NLP) technology has been rapidly evolving in recent years. Being the pioneer in this area, we are motivated to undertake this project to explore the fusion of VR and generative AI in improving the existing soft skills training applications.

1.2 Project Scope

The scope of this project is confined to developing an immersive soft skills training application by leveraging Large Language Models (LLMs) and Virtual Reality (VR). The final deliverable will be a Unity application which can be run in desktop mode that provides visuals and sounds of LLM-controlled avatars, as well as a user interface that takes trainees' speech as input. The client application is backed by a Python FastAPI server, comprising the novel multiagent architecture that interacts with LLMs to simulate believable agents' behaviour. This Unity application is accessible to all trainees who are desperate to practice a variety of soft skills in an immersive environment. The trainees can complete a short scenario-based simulation practice and retrieve comprehensive feedback. On the other hand, a web interface can be accessed by course creators, potentially university soft skills lecturers, or corporate trainers to design customized simulation practices and monitor the trainees' performance over time.

1.3 Project Objectives

In this project, we aim to develop an immersive soft skills training application using LLMs and VR. The details of each sub-objective are explained below:

I. To develop an LLM-powered Simulation Module for simulating real-world scenarios

The project aims to develop a simulation module that can simulate the real-world social scenario based on the course creators' input. The agent avatars behave according to trainers' requirements, and trainees seamlessly interact with the agents just like having natural conversations with people in a real-world scenario.

Within this simulation module, we intend to design a partially centralized multiagent architecture which is supported by LLMs. The central system coarsely defines the plans for each avatar agent, and each agent has a degree of autonomy to finely adjust and execute the plan. Our proposed architecture improves agents' cognitive abilities while ensuring time-efficiency and cost-effectiveness.

On the other hand, an action projection layer, which supports different kinds of modalities transformation, is embedded into the simulation module. For instance, LLM-generated text can be converted to visual and auditory modalities, and, conversely, trainees' speech can be converted into text with nonverbal information tagged. This projection layer enables the seamless integration of LLMs into VR environments.

II. To develop an LLM-powered Scenario Creation Module for creating simulation practices efficiently

We intend to develop a scenario creation module that can assist the soft skills experts to create scenario-based simulation practices with minimal effort. Compared to most existing applications, our module does not require trainers to enter each possible outcome of a scenario, instead, they are only required to enter basic details such as the duration of a simulation, the brief description of the scenario and desired learning objectives. To ensure the module aligns with the user requirement, we collaborate with UTAR soft skills lecturers to enable continuous system refinement.

III. To develop an LLM-powered Evaluation Module for improving feedback comprehensiveness

We aim to develop an evaluation module that not only evaluates trainees' performance based on quantitative data but also combines qualitative information according to the learning outcome. Quantitative data includes speech rate, speech loudness and speech energy which usually can be evaluated using heuristic rules or machine learning models. Unfortunately, these evaluations lack context-awareness even though valuable insights are provided. Hence, LLMs are integrated to assess the simulation transcription based on the list of evaluation criteria defined by course creators. Overall, this module gives a scoring breakdown and personalized feedback to help trainees improve themselves over time.

IV. To analyse the LLM-generated utterance and qualitative feedback for ensuring cost-effectiveness

LLMs are used to perform different tasks such as cognitive reflection, action execution and qualitative evaluation. To ensure LLMs generate expected behaviour and accurate feedback in terms of soft skills training, we will examine the LLMs' ability to align their responses with human input requirements by conducting extensive experiments. Ultimately, we aim to show that our LLM-based application is cost-effective at scale despite the computational cost of LLMs.

1.4 Contributions

Our project has 4 contributions. Firstly, a VR soft skills training application system that adopts an LLM-based multiagent framework is developed. While existing literature focuses on advancements within LLM-based multiagent frameworks or VR technologies separately, our project pioneers the integration of these two areas. For instance, while the proposed agent architecture in this paper [8] demonstrated LLM agents' believable behaviours, it falls short in meeting real-time constraints, which is a critical requirement in VR. On the other hand, while VR soft skills applications that utilise LLMs exist, such as Virtual Speech [3], the underlying language model is often poorly integrated.

Secondly, the proposed multiagent architecture has improved cognitive reflection capabilities and time efficiency compared to most existing research. This is realised by improving the agent architecture in [8] to fit the context of soft skills training applications. To achieve cost-effectiveness, our architecture strikes a balance between maximising LLM output quality and constraining the frequency of LLM prompts. Besides, LLMs are evaluated to ensure the selection of the most cost-effective LLMs.

Other than that, we combine quantitative and qualitative data when evaluating trainees' performance. Existing applications only focus on quantitative data when providing feedback, even though Virtual Speech [3] has integrated LLMs, the qualitative evaluation is not well designed. The fusion of two different types of trainees' data requires an understanding of the pros and cons of LLMs, for instance, LLMs specialise in language reasoning rather than complex mathematical reasoning. In our project, we streamline the process to ensure both qualitative and quantitative data are evaluated.

Lastly, the proposed application will be beneficial to university students and corporate workers who are required to gain soft skills to increase workplace competencies. With a more realistic immersive environment and comprehensive evaluation, a greater number of trainees can now access high-quality soft skills training modules. This can prepare lifelong trainees for their career success, enhance their employability, and promote Malaysian economic growth. It is noteworthy that our simulation practice creator will attract soft skills course creators to create various scenarios, fostering continuous improvement of our system and ensuring its long-term sustainability.

1.5 Background Information

1.5.1 Soft Skills

Soft skills refer to the ability of an individual to communicate, collaborate and interact with others in various settings, including workplaces, universities, and social environments [9]. According to UTAR DSSC, soft skills can be classified into 10 categories, which include Communication and Language Skills, Critical and Analytical Thinking, and Leadership Skills [10].

While university students typically acquire hard skills throughout their studies, it is not enough to enhance employability without practising soft skills. As Malaysia undergoes a significant transition towards a skills-based economy, Koo [1] emphasised that maintaining competitiveness requires proficiency in both soft and hard skills. As such, many jobseekers in Southeast Asia strive to gain soft skills to improve themselves even during economic downturns [1]. Through the LinkedIn Learning platform, it was observed that out of the top 10 trending courses, 7 were centred around soft skills, including public speaking, communication foundations and project management in 2022 [1].

1.5.2 Virtual Reality

According to [11], Virtual Reality is known as the creation of interactive 3-dimensional (3D) worlds by using computer technology, where objects in the virtual worlds evoke a feeling of spatial presence. The essence of VR is the realistic cognitive experience which makes users feel like they are interacting in the real world.

In this project, we study the ways to improve the cognitive experience of existing VR applications by adopting generative AI, which can enhance virtual avatars' cognitive behaviour. To be classified as a VR application, this project will: 1) adopt computer technology; 2) generate 3D environments with 3D avatars; and 3) enhance the feeling of "spatial presence" by improving avatars' cognitive behaviour. As such, a VR application can be run on either a VR headset or a personal computer.

1.5.3 Large Language Models

LLMs are deep learning models that have strong text understanding and text generation capabilities [12]. They function by progressively estimating the probability of the next tokens, which are chunks of characters. Unlike traditional language models, LLMs are trained on massive datasets and possess a large number of parameters that allow them to understand and generate longer sentences. However, training and using LLMs require high computational power. While most proprietary models have limited customization options, the developers can use the extensive knowledge and language reasoning power of LLMs to create their applications. To do so, they need to design effective prompts that elicit the optimal outputs to address specific business needs. This process is called prompt engineering recently.

In this project, we are proud to practice prompt engineering, ensuring the prompts are optimised. Besides, a novel multiagent architecture that maximises LLMs' capabilities will be introduced.

1.6 Report Organization

This report contains 7 chapters. In this chapter, we highlight the problem statement, project scope and objectives, discussion of impact and contributions, as well as essential background information. In Chapter 2, we discuss past research that worked on NLP in roleplay and text analysis, as well as existing similar projects with strengths and weaknesses analysed. In Chapter 3, we explain the system architecture, methodologies, tools used and the project plan. In Chapter 4, we discuss the system design, including detailed architecture of each system module. Subsequently, Chapter 5 showcases the system implementation results along with system screenshots. Next, Chapter 6 defines performance metrics and displays the system testing results along with objective evaluation. Finally, Chapter 7 concludes our findings and highlights the overall novelty of our project.

Chapter 2 Literature Review

2.1 Natural Language Processing in Roleplaying

In this section, several recent NLP research studies are presented. Before the rise of ChatGPT, the area of leveraging deep learning methods for generating human-like conversations has been widely discussed in recent years. To create realistic virtual avatars, it is imperative to investigate the NLP capabilities in roleplaying.

2.1.1 Dungeons and Dragons as a Dialog Challenge for Artificial Intelligence

In this study [13], the text-based adventure game, Dungeons and Dragons (D&D) served as an ideal arena for a thorough exploration of the language generation, language understanding, and strategic planning capabilities of LLMs.

The study proposed a novel dataset constructed by scraping data from an online forum. Through heuristic labelling, the dataset accurately mirrors realistic conversations, considers historical conversations, involves multiple participants, and includes storytelling elements. To examine the robustness of their datasets, Google LaMDA, the chosen LLM was finetuned against the proposed dataset to perform two tasks, namely next utterance prediction and game state tracking. The first task was to predict the next utterance based on the input conversation, while the second task focused on predicting the ongoing game state based on the given conversation.

The performance was evaluated by a group of professional raters. Next utterance prediction was assessed based on the generated conversational history, with ratings assigned according to the coherence, specificity, and interestingness of the models. On the other hand, game state tracking was evaluated based on the accuracy in predicting the current turn's state variables. As a result, finetuned LaMDA yielded a better result than conventional dialogue systems in terms of the next utterance prediction. However, despite finetuning efforts, game state tracking results still exhibited low accuracy. The authors suggested that exploring alternative models might be necessary.

The contribution of this study is the novel D&D dataset as described. Although this paper showcased the capability of LLMs to generate more natural responses within a conversation, the result is currently limited to the D&D context. This implies the need for scraping, annotating and finetuning efforts to enable LLMs to roleplay effectively in different contexts. While the existing limitations may be attributed to the inherent characteristics of the chosen LLM, it is noteworthy that recent commercial LLMs such as GPT-3.5 [14], GPT-4 [14] and Gemini Pro [15] have significant improvement in language reasoning capabilities. Such improvements can be attributed to the fact that these foundational LLMs are trained on larger datasets and possess a larger number of parameters. Hence, in our project, we opt for GPT-3.5-Turbo as the foundational models. Then, prompts will be designed in a way that adapts to different roleplaying contexts, ensuring the extensibility of our architecture.

2.1.2 Dialogue in the Wild: Learning from a Deployed Role-Playing Game with Humans and Bots

This paper [16] argued the limitations of existing data-driven NLP research which include the lack of natural and realistic characteristics in crowdsourced data, and the inability to learn from historical experiences within static datasets. To overcome these shortcomings, this paper proposed a framework that allows iterative training of NLP models through data provided by intrinsically motivated human participants. The approach involved the development of a text-based role-playing game with a fantasy theme.

The methodology was a cyclic process of retraining and deploying the NLP models, originating from base models which were trained using crowdsourced data. The designed role-playing game was then deployed through Facebook Messenger, and it was advertised via Facebook as well. The data collected from players was sophistically processed, with the quality of players' responses evaluated using a role-playing score model. Throughout the process, several hypotheses were made, for instance, the positive correlation between human engagement levels and response quality, and the direct influence of model quality on human data quality.

To evaluate the proposed framework, the performance metrics such as accuracy were assessed in every round. Additionally, cost learning curves were plotted to analyse the long-term feasibility of the model, considering the cost of online advertisement for attracting players. The findings indicated the initial feasibility of the proposed framework; however, the long-term feasibility of this framework remains unclear. The proposed framework still heavily relies on training data, signalling a potential challenge in adapting the system to different contexts. Nevertheless, the iterative training concept is commendable. In our project, while we use LLMs as underlying models without regular finetuning, continuous data retrieved from users are still useful for enhancing the framework and iteratively refining the prompts for LLMs over time.

2.1.3 LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation

The study [17] attempted to enhance chatbots in the field of mental health by utilizing one of the LLMs, GPT-3.5 Turbo to simulate interactions between doctors and patients. The LLM chatbot development and evaluation frameworks were proposed to facilitate the entire development process.

As shown in Figure 1, the development process comprises three phases, wherein prompt engineering was conducted iteratively. Guided by continuous input from professionals, the criteria for high-quality doctor and patient chatbots were established. Doctor chatbots were designed to exhibit comprehensive medical knowledge, in-depth questioning, and empathetic responses. Conversely, defining the quality of patient chatbots was challenging due to the absence of general guidelines. Hence, the initial requirement for patient chatbots was to ensure their honesty.

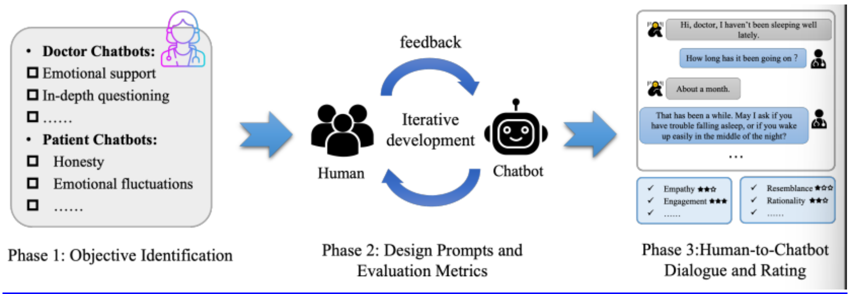


Figure 1 Iterative development of chatbots [17]

Throughout the doctor chatbot development, limitations were identified in GPT-3.5's ability to generate in-depth questions and empathetic responses. The limitations were mitigated by embedding relevant clinical domain examples into the context. For patient chatbots, critical patient behaviours such as "depressed emotions", "difficulty in expressing symptoms" and "resistance to seeking help" were incorporated into the prompts as suggested by psychiatrists. While developing such behaviours, it became obvious that GPT-3.5 could forget their behaviours. Hence, the issue was managed by temporarily appending reminders within the most recent prompts to reinforce the expected characteristics.

The implementation was done by using the standard approach offered by OpenAI, which was combining system, user, and assistant messages to generate responses. The system message of each chatbot consisted of roleplaying information and a series of instructions that the chatbots were required to follow throughout the conversations. For instance, the doctor chatbots were required to conduct diagnosis and ask one in-depth question at a time. To find out the best prompts, some of the instructions in a prompt were omitted during evaluation, and different prompt variants were compared against each other.

The chatbots' performance was evaluated by depressed patients and psychiatrists. The doctor chatbots' evaluation focused on user experience fluency, empathy, professionalism, and engagement. Meanwhile, patient chatbots were evaluated based on resemblance and rationality. Compared to human doctors, doctor chatbots struggled to evenly distribute the questions across various symptoms. Besides, doctor chatbots were also less comprehensive in probing specific symptom conditions. In the case of patient chatbots, it was noticed that constantly appending reminders to the prompts enhanced its retention of role-playing behaviour. However, this approach led to an increased likelihood of these chatbots forgetting their actual symptoms.

In summary, the paper presented the feasibility of LLM chatbots to simulate human interactions within the mental health domain. Nevertheless, the developers had limited control over the generated responses due to the reliance on the pre-trained dataset used to train GPT-3.5. Although the system message was properly designed for roleplaying, multiple instructions in a prompt occasionally challenged GPT-3.5, an attention-based model. Besides, the generic mechanism resulted in lengthy prompts and reduced output quality as the conversation proceeded. In our project, we effectively mitigate this limitation by adopting the Retrieval Augmented Generation (RAG) framework [18]. This involves periodic summarization of the conversation history and storing all additional information in a vector database. The querying process selectively retrieves the most relevant information, and subsequently, the queried result is used to prompt the LLM. This approach not only improves the output quality but also optimizes the input token consumption.

2.1.4 Generative Agents: Interactive Simulacra of Human Behaviour

The study [8] presented a novel generative agent architecture to showcase the possibility of LLM-powered agents in simulating believable human behaviours. As a proof of concept, a simulation involving 25 agents with different characteristics was conducted in a sandbox environment. In the simulation, GPT-3.5-Turbo was the chosen LLM.

The agent architecture is shown in Figure 2. When a new event occurs in the environment, the observation is stored in a memory database, and subsequently, a memory retrieval algorithm selects and extracts the most relevant fragment of memory from the memory database. This extracted memory is then used to prompt the LLM to perform reflection and planning. LLM plays a role in generating thoughts and plans according to the observed event. Eventually, the agent executes the plan by interacting with the environment. It is noteworthy that the agent may choose to continue or change the existing plan during the plan generation process.

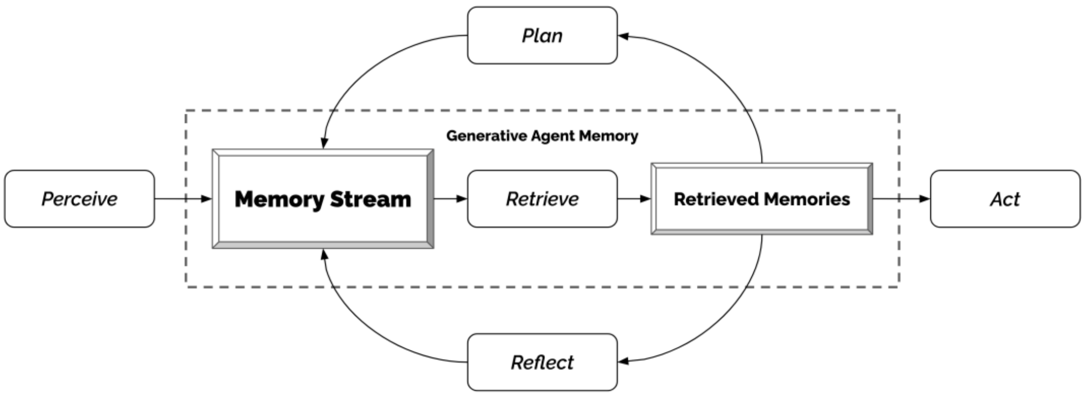


Figure 2 Generative Agent Architecture [8]

The architecture was evaluated through a holistic procedure. This includes "interviewing" the agents to verify whether their characteristics, plans and responses align with their characteristics and memories. Then, human evaluators were invited to rank the believability of the simulated agents across various architectural configurations. As a result, agents with full architecture generated the most believable behaviour.

Overall, there are several contributions of this research. Firstly, the generative agents can adapt their behaviour in response to environmental changes and influence the environment. Notably, these adaptations occurred without the need to hardcoding their behaviours. Besides, the novel agent architecture allowed efficient memory retrieval, which can effectively mitigate known issues associated with LLMs, such as hallucination.

The proposed novel architecture demonstrates commendable potential for application across different problem domains. However, agents can occasionally generate inconsistent outputs, such as hallucinations when memory retrieval fails. Besides, the instruction tuning effect of the underlying GPT-3.5-Turbo model tends to make the agent behave too formal and cooperative, limiting its potential for generating human-like behaviour. Finally, the architecture falls short in terms of cost-effectiveness despite its potential for generating believable human behaviour.

In the context of soft skills training, the agents should execute behaviours and characteristics that match the course creators' expectations. Hence, our project improves this architecture by ensuring the reflection process aligns with the expected cognitive behaviour. Besides, we simplify the planning module as extensive planning is not necessary for our use case, this greatly reduces the computational cost.

2.2 Natural Language Processing in Text Analysis

In this section, we present the existing studies which have explored the capabilities of NLP in analysing text.

2.2.1 Use of Natural Language Processing (NLP) Tools to Assess Digital Literacy Skills

In this paper [19], they proposed NLP tools that could assess one of the soft skills, namely digital literacy skills to assist teachers in providing timely, reliable, and personalized feedback. The evaluation of such tools was done by experimenting with them against a total of 286 students.

The experiment was initiated by 4 different groups of students which were the experimental group with and without pretest, as well as the control group with and without pretest. As shown in Figure 3, 1st and 3rd groups of students underwent a pre-test which focused on vocabulary, critical reading, and writing. Then, 1st and 2nd group of students went through three different types of training activities including discussions and seminars. In the end, all students underwent a post-test to examine their skills again. The results of the pre-test and post-test were then fed into the NLP model to analyse their skills.

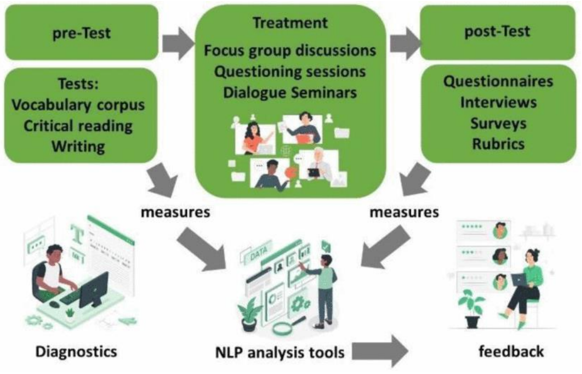


Figure 3 Experiment for NLP Analysis Tools Development [19]

To perform NLP analysis, student content in the form of video essays or podcasts was first transcribed into text using speech-to-text tools. After tokenizing the text, preliminary measurement was done on the frequency of use of the words. Through this step, information such as fluency and structural complexity of the text were extracted. Then, the level of proficiency was further categorized based on length, quality, and complexity. Lastly, semantic information was also extracted with the help of lexical analysis to help the model understand text meaning.

The result incorporated vocabulary, syntactic and semantic analysis. It showed that student groups which attended the training activities yielded significant improvement in communication skills and cognition maturity compared to the control group.

Ultimately, the contribution of this study includes the integration of semantic information into the NLP tools. As part of the future work, the research team claimed to extend the functionality by adding sentiment analysis and gesture identification, to analyse the performance of nonverbal communication as well.

However, even though semantic information was incorporated, traditional lexical analysis has limited reasoning capabilities compared to LLMs, as LLMs have demonstrated their proficiency in processing common sense semantics [20]. Nevertheless, this study underscores the importance of incorporating context understanding to create a more comprehensive soft skills evaluation system.

2.2.2 Is ChatGPT Equipped with Emotional Dialogue Capabilities?

The study [21] attempted to explore the emotional dialogue capabilities of GPT-3.5 Turbo based on 2 main aspects, namely understanding capability and generative capability. The author benchmarked ChatGPT results with different state-of-the-art (SOTA) emotion recognition models, however, no matter whether ChatGPT was used in zero-shot or few-shot manners, ChatGPT generally had a performance gap between these advanced models on Emotion Recognition, Emotion Cause Recognition and Act Classification task. This is because all baseline models were taking a supervised approach, where ChatGPT performed sentiment analysis based on its interpretation and standards. The author highlighted that not every question has definitive answers, for instance, a dialogue containing the statement "Thank you very much" was annotated as neutral, while ChatGPT interpreted it as "happiness".

Generated responses of various models were evaluated both through automatic ways and by human experts. In terms of generative capability, ChatGPT could generate longer, and complex responses compared to other models. However, the author noticed that ChatGPT was often too eager to give advice rather than empathetical messages.

As a future direction, the author highlighted the importance of enhancing ChatGPT capability to a personalized level. It should be done by designing more useful prompts, also known as prompt engineering, to enhance ChatGPT's ability to generate empathy. The author also highlighted the limitations of this study, which include the model, ChatGPT selected may not fully represent all LLM models.

Overall, this paper well analysed the ChatGPT's capability of emotional dialogue, however, the ChatGPT used for this study did not perform prompt optimization to better personalize the generated responses. Without a comprehensive architectural design, the comparison of ChatGPT models with other well-trained models may not fully explain the capability of ChatGPT. It was unfair to compare ChatGPT against other NLP models that may have been fine-tuned for specific datasets, while ChatGPT relied solely on pre-trained data. To address this limitation, our project adopts a human-centred design approach. For instance, course creators can define the evaluation criteria for simulation practice, allowing LLMs to have a clear direction of human expectation rather than autonomously analysing and classifying trainee outcomes themselves. This approach fully utilises LLMs' complex reasoning capabilities.

2.3 Soft Skills Training Application Review

2.3.1 Virtual Speech

Virtual Speech is a VR soft skills training platform which was founded in the UK in 2016 [3]. Their clients are universities, Fortune 500 corporations and 370k users spanning more than 125 countries. Traditionally, individuals train their public speaking skills in front of the mirror, however, this experience is unable to replicate the experience of addressing a live audience. Hence, the platform was initially designed to simulate a real-world public speaking environment using VR technology. With this successful concept, the platform has expanded its VR training activities to sales pitches, interview sessions and other activities.

Virtual Speech is accessible through its website. Trainees can access a self-paced learning module that offers reading materials, scenario-based practices and quizzes.

Strength

a. **VR-enabled Practice Room**

To simulate the real-world scenario, Virtual Speech provides a series of practising environments which includes live conference, office meeting, interview, classroom, eye contact training and impromptu speech training.

All these scenarios are VR-ready. This feature can be accessed either through their website or by installing their application from the VR Oculus Appstore. Trainees can opt for either 2D animated virtual characters or 3D avatar characters for each scene.

For rehearsal purposes, trainees can upload their presentation notes or use the sample notes available within the system. As shown in Figures 4 and 5, trainees can access and control the display of these notes while virtually presenting to virtual avatars. This feature empowers trainees to practice their presentations multiple times, ensuring they are well-prepared for their upcoming meetings or presentations.

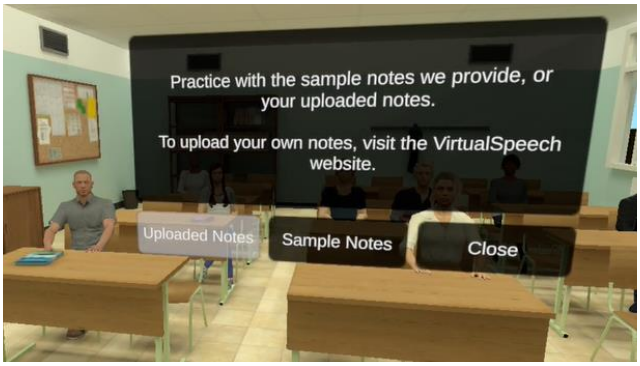


Figure 4 Presentation Scene - Virtual Speech

Figure 5 Presentation with Scripts - Virtual Speech

b. Quantitative Analysis

At the end of a virtual practice, the system generates a graphical report analysing various quantitative aspects of speech and gestures. As shown in Figure 6, the evaluation metrics include speaking pace, volume, usage of filler words, audibility, and eye contact. The focus of these attributes is to improve trainees' nonverbal communication skills without looking into the context of the speech.

Figure 6 Quantitative Analysis - Virtual Speech

Weakness

a. Lack of contextual understanding in certain scenarios

Besides public speaking, some scenarios such as interviews and sales pitches may involve direct interaction with virtual avatars. However, as of 27 October 2023, the avatars in these practice rooms were only capable of asking predefined questions to trainees, and they could not understand or respond to trainees' answers. An example of the virtual interview is shown in Figure 7, where trainees were required to press the "start" button to initiate the avatar's questioning process. In our system, we ensure all the scenarios are driven by LLMs and course creators can customize avatars' behaviour.

Figure 7 Virtual Interview - Virtual Speech

b. Limited ChatGPT Integration Capability

As of 27 October 2023, the platform has integrated the ChatGPT API into some scenario-based practices. These scenarios, as shown in Figure 8, provided trainees with opportunities to interact with a ChatGPT-powered avatar. However, these interactions were limited to 1-to-1 conversations, and there were time constraints on the duration of the interactions. As tested by us, the practice session would be terminated by the system after trainees responded to around 5 questions from ChatGPT. This setting is generally understandable when the system was developed using generic chat completion functions offered by OpenAI. Without performing prompt optimisation, prompts which consist of the conversation history will become increasingly lengthy. Hence, the system may need to limit user prompts in such cases to prevent excessive API charges resulting from high token consumption.

In our project, we present a VR application that enables the simulation of multiple avatars, powered by our novel multiagent architecture as the backend. Besides, our agent architecture ensures the prompts do not exceed a specified token limit by regularly summarising the conversation history.

Figure 8 ChatGPT Roleplaying - Virtual Speech

2.3.2 Mursion

Mursion [4] is another VR soft skill training provider in the market. It is a US company which strives to offer corporate training with education technology. Training contents including leadership, diversity equity and inclusion, and customer service are offered.

Strength

a. Human-controlled avatars

Instead of creating AI-powered avatars, Mursion uses human-controlled avatars to interact with trainees. Since the avatars are controlled by humans, the application can handle situations involving multiple avatars. This is beneficial for simulating real-world scenarios that require a lot of interactions, for instance, teaching kids as shown in Figure 8. Mursion claimed that such human-controlled avatars ensure a safe and risk-free virtual environment.

Figure 9 A teacher-student Roleplaying Session – Mursion [22]

Weakness

a. Potential Scalability Issues

The avatars can be realistic; however, the need for human control downgrades the scalability of the system. While human-controlled avatars ensure the safety of the environment, the system will not be always available as it depends on the presence of Mursion workers. As the scene can be repetitive, utilising LLMs in creating virtual avatars can be beneficial. While periodic human reviews of AI-generated responses may be necessary to ensure the system's safety, it allows the application to be scalable with minimal human effort.

2.3.3 STRIVR

This is a VR training platform which utilises VR to develop hard skills and soft skills for corporate trainees [5]. Their in-house content team is responsible for customising the practice scenarios specifically based on their client's requirements. The VR environment can be a recorded video scene. For each simulation scenario, trainees are offered multiple choices to continue the roleplaying session.

Strength

a. Immersive Learning

For the soft skills training module, they demonstrate different scenarios such as customer service, sales training and new hire onboarding. The AI-powered VR platform simulates realistic scenarios like dealing with unhappy customers and other complex situations, encouraging trainees to practice de-escalation methods and empathy building. Under such controlled environments, trainees can confidently interact with the scenario, unlike traditional role-play activities.

b. VR-based Assessments

To help hiring managers better assess candidates' skills, STRIVR offers VR-based assessments where users are given a designed task in an immersive environment. Based on their own AI models, they can predict the users' performance by combining decision-related data and immersive attention data gathered during the test. While the AI analysis may not fully reflect the candidates' skills, this offers useful insight for hiring managers to better understand the candidates or employees.

Weakness

a. **Decision-Tree-Based Storyline**

However, based on the product demonstration [23], most scenario creations rely heavily on human inputs. The scenarios given are pre-recorded, and trainees are given predetermined options rather than being encouraged to make their own decisions. This is typically the limitation of the traditional roleplaying games approach, where the backend is driven by decision-tree or rule-based methods. When predefined options are provided, this degrades the authenticity of the simulated environment as trainees tend to select the options that appear safe or correct within the context. Nevertheless, we believe that the integration of generative AI can significantly improve this limitation by adapting the storyline according to the spontaneous reactions made by trainees.

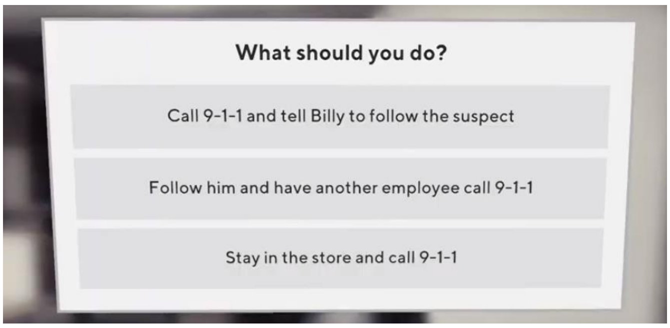


Figure 10 Decision-Tree-Based Simulation [23]

2.3.4 Orai

Orai [7] is a mobile application that is targeted to train trainees' communication skills by practising verbally. Unlike the applications mentioned above, it does not provide an immersive experience. Nevertheless, it provides a quantitative analysis to trainees regarding their speech. With the convenience of mobility, users can practice anywhere along with their phone.

Strength

a. Self-paced learning

Like most training platforms, Orai provides a self-paced learning module which consists of different chapters, to help trainees master communication skills. Figure 10 shows the learning page, namely 'Journey', where trainees need to continue the earlier section first before proceeding to the next section. The whole learning module is a mix of learning content, quizzes, and guided practice.

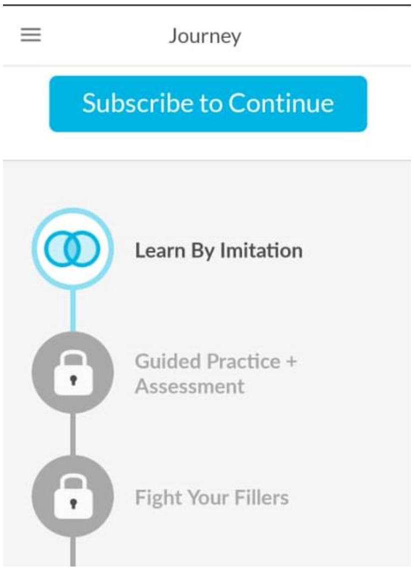


Figure 11 Self-paced learning – Orai

b. Guided practice with Quantitative Analysis

In the guided practice, trainees are given a topic, and they need to give a short speech within minutes. The application records the trainees' audio and quantitatively analyses the trainees' energy, speech tone and speech rate after the session. As shown in Figure

12, the app transcribes trainees' speech, with proper annotation to inform trainees when certain parts are too speedy or too slow.

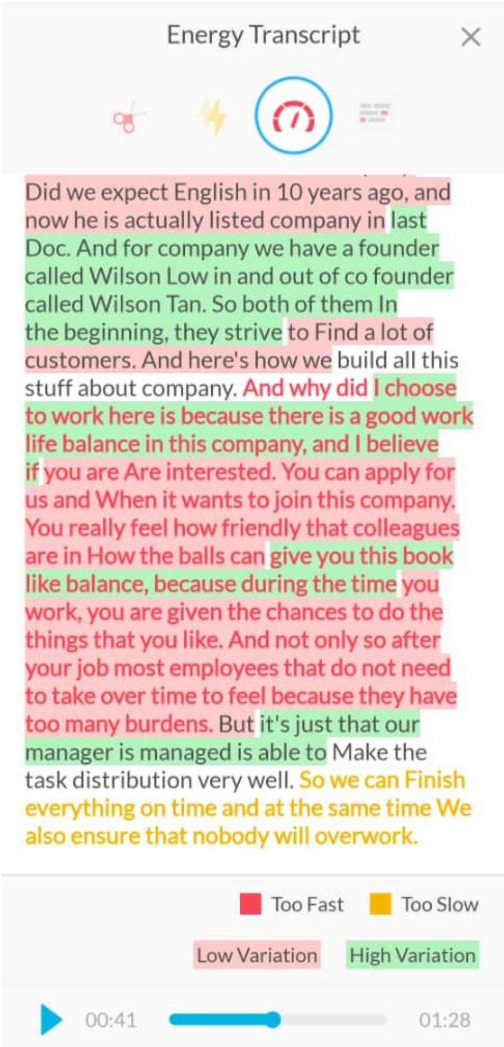


Figure 12 Energy Transcript - Orai

At the end, the app outputs a total score of the practice. As shown in Figure 13, the trainee is advised to improve on energy, pace, and conciseness.

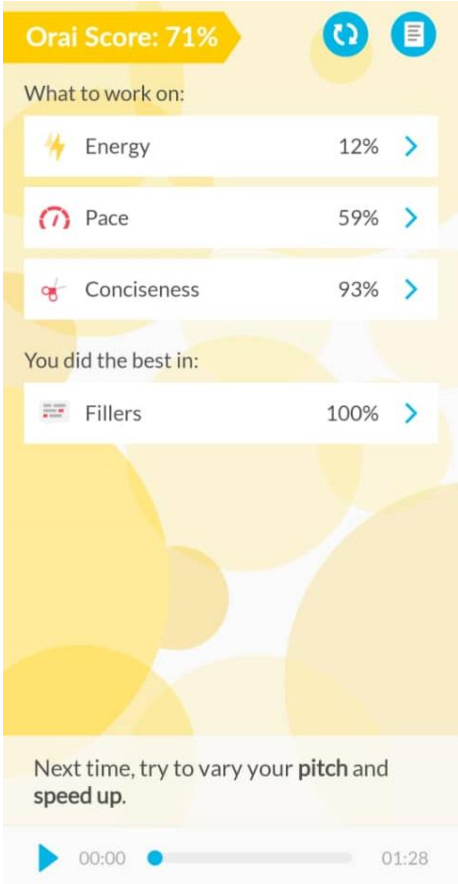


Figure 13 Overall Orai Score

Weakness

a. Lack of contextual understanding

The performance analysis provided only relies on quantitative data such as speech rate and speech tone. However, it does not analyse the speech according to contextual information. Even though the application transcribed the users' speech, the analysis provided has no connection to the content's meaning. When contextual information is not considered, trainees are prone to overly focus on quantitative evaluation without paying attention to their speech content.

b. Lack of realistic environment

Unlike other VR platforms, this application did not provide any simulated scenes for trainees. To undergo the exercises offered, trainees simply need to speak to the phone without facing any virtual audience. Consequently, this approach failed to simulate the feelings of anxiety and nervousness that real-life situations may evoke. As such, it has limited capability to train the trainees' communication skills.

2.4 Comparison of all the models and solutions

Table 1 Comparison of all NLP Models

Paper	**NLP in Roleplaying**			
	Dungeons and Dragons as a Dialog Challenge for Artificial Intelligence	Dialogue in the Wild: Learning from a Deployed Role-Playing Game with Humans and Bots	LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation	Generative Agents: Interactive Simulacra of Human Behavior
Strength	- Proposed D&D dataset with improved next utterance prediction result.	- Iterative training of NLP models is cost-effective.	- Iterative development framework for prompt engineering	- novel agent architecture that generates believable human behaviour
Weakness	- game state tracking shows low accuracy - Limited adaptability	- limited adaptability	- Limited control over responses generated by GPT-3.5-Turbo	- occasional inconsistent outputs - not cost-effective

Table 2 Comparison of all NLP Models (cont.)

Paper	**NLP in Text Analysis**	
	Use of Natural Language Processing (NLP) Tools to Assess Digital Literacy Skills	Is ChatGPT Equipped with Emotional Dialogue Capabilities?
Strength	- integration of semantic information into the NLP tools	- well analyse the ChatGPT's capability
Weakness	- Limited context understanding capabilities	- lack comprehensiveness when benchmarking ChatGPT with other NLP models

Table 3 Comparison of all Soft Skills Training Applications

Paper	**Soft Skills Training Applications**			
	Virtual Speech	Mursion	STRIVR	Orai
Strength	- VR-enabled practice room - Quantitative Analysis of trainees' performance	- Human-controlled avatars with high safety	- VR-based training and assessment	- self-paced learning module - guided practice with quantitative analysis
Weakness	- Lack of contextual understanding in some scenarios - Limited ChatGPT integration capability	- potential scalability issues due to the need for human control	- decision-tree-based storyline degrades authenticity	- lack of context understanding when performing analysis - lack of realistic environment

2.5 Critical Remarks

Firstly, previous studies have limited exploration of roleplaying using LLMs. Existing models either lack adaptability or produce outputs that do not align with the users' expectations.

Secondly, the exploration of using LLMs for text analysis is a relatively new approach. Existing NLP models fall short of providing comprehensive feedback due to limited context-understanding capabilities.

Finally, most commercial soft skills training applications still have significant room for improvement, as some lack contextual understanding in giving qualitative feedback, and the immersive experience can be further improved with more human-like avatars.

In our project, a novel LLM-powered multiagent architecture is presented with improved cognitive behaviour generation. Besides, we adopt a human-centred approach to allow course creators to define learning outcomes to examine trainees' soft skills. This ensures that the system provides high-quality feedback without compromising scalability. Lastly, we stand as a leading project that incorporates novel generative AI frameworks with VR environments, highlighting the potential for commercialisation.

Chapter 3 System Methodology/Approach

In this chapter, we discuss the system methodology and approach in advance.

3.1 System Design Diagram

3.1.1 System Architecture Diagram

The proposed approach is an immersive soft skills training application which utilizes Large Language Models and Virtual Reality. It can be used by trainers to design a variety of scenario-based simulation practices to train the trainees. The proposed system adopts a client-server architecture, and Figure 14 is a system architecture diagram that shows a high-level view of this system. The system is composed of a server application and two client applications: a Unity application and a web application.

As shown in Figure 14, trainees can interact with multiple agent avatars through the Unity application. The simulation practice begins when the Unity application establishes a WebSocket connection with the server. Once the simulation session starts, the central system thread and agent avatars threads, both supported by LLMs, start to perform their routine tasks. The central system thread coarsely defines the action plans for each avatar. Then, agent avatar threads execute the actions which can be either generating speech or displaying gestures. To convert the text generated by LLMs into visuals, the text is then passed to the action projection layer. The projection layer is an extensible module that handles transformation of different modalities, for instance audio and body gestures into text, and vice versa. The role of Unity application is to transmit input from trainees, which is currently, limited to speech, to the action projection layer at the server, and conversely, receive and present avatars' speech and gestures from the action projection layer.

On the other hand, the web application is crucial for trainers to define a practice scenario and for both trainers and trainees to retrieve the trainees' performance evaluation. The trainers input the scenario information and learning goal, and with the assistance of LLMs, a customised simulation configuration that includes the details of virtual avatars is generated. Then, trainees who have access to the training can start the simulation practice using the Unity application as explained above. Upon completion of the practice, the transcript including the dialogue history and speech meta-information is stored and passed to the evaluation module. The LLM-powered evaluation module analyses the trainee's performance from both quantitative and qualitative aspects. Finally, the evaluation result can be obtained from web application by trainers and trainees.

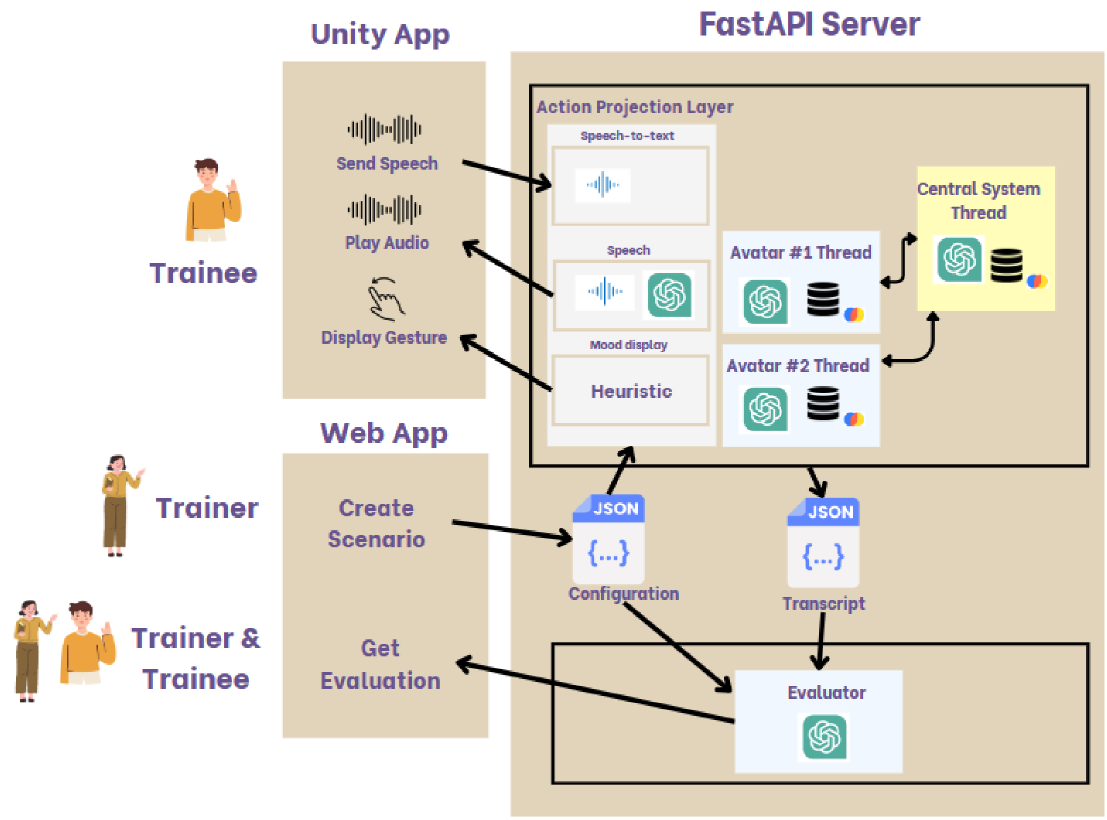


Figure 14 System Architecture Diagram

A system design is then clearly defined based on the architecture diagram above in Chapter 4.

3.4 Timeline

Table 4 Project Timeline for Report 1

Project Task	**Project Week**					
Activity\Week	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**
Problem Statement and Project Scope Refinement						
Analysing Third-Party Services						
System Design and Prototyping						
Finalising System Design Specification						
Finalizing Report 1						
Presenting Report 1 with Prototype Demonstration						

Table 5 Project Timeline for Report 2

Project Task	**Project Week**												
Activity\Week	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Week 9**	**Week 10**	**Week 11**	**Week 12**	**Week 13**
Resolving Project 1 Limitation													
Project 1 Code Refactoring													
Implementing side features													
System Testing													
Finalizing Project Deliverable and Report 2													
Presenting Report 2 with Application Demonstration													

Chapter 4 System Design

4.1 System Block Diagram

As explained in Chapter 3, our system adopts client-server architecture, with the main functionalities of scenario practice creation, real-time simulation and evaluation. Hence, our system contains a scenario creation module, a simulation module and an evaluation module, as shown in Figure 15. The whole training process starts with a trainer creating a new scenario practice, for instance, an interview session. Then, the configuration of the simulation, which is a collaborative effort of both the trainer and LLMs, is stored. The simulation module takes the configuration file as the input, and the trainee can start to engage with the virtual avatars and the environment in the simulation. Once the session is completed, the transcript is saved, and the evaluation module starts to evaluate the trainee's performance. Finally, both the trainer and the trainee can have the access of result dashboards, which states the strength and weaknesses of the trainee in the simulation.

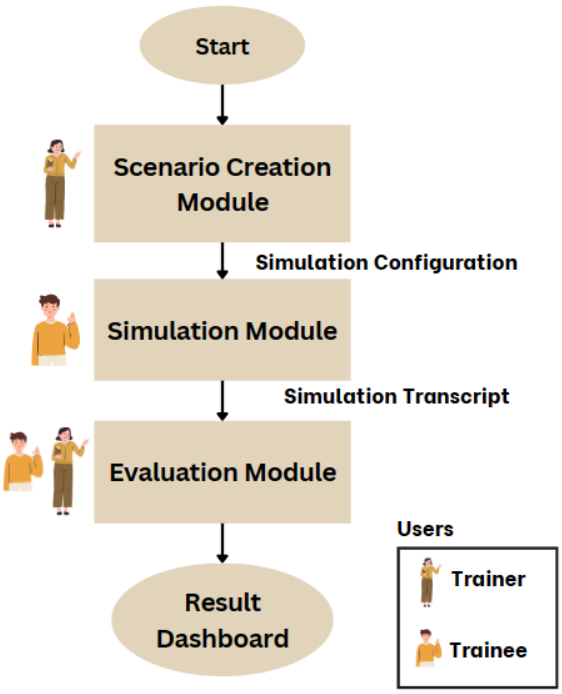


Figure 15 System Block Diagram

4.2 System Components Specifications

4.2.1 Scenario Creation Module

This module serves as the entry point for a VR soft skills training session. Trainers can interactively insert the simulation practice details with AI generation feature.

As shown in Figure 16, our system first displays a series of example scenarios to the trainer. The trainer can choose to use the given template or type their own one from scratch. The required information includes scenario name, a short scenario description, number of avatars and simulation duration. Subsequently, the scenario details are passed to the scenario generator to generate the details of each avatar. The scenario description is part of the prompts for the LLM to maintain the storyline in a simulation. Notably, prompts for the LLM must be aligned with certain implicit formats to ensure optimal performance. The defined format is to keep the scenario description short, with at least all the avatars' name, including trainees' name to be told. However, to prioritize user experience, trainers should not be burdened with these concerns. Hence, our scenario generator, supported by GPT-3.5, will help refine the scenario description when generating the avatar details. Figure 17 shows our prompt template for the scenario generator.

Once the trainee roleplaying and avatar details are generated, they are displayed to the trainers. Figure 18 shows the relationship between a scenario, an avatar and a controllable avatar. In our context, trainees are roleplaying as a controllable avatar, hence, the controllable avatar shares the same attributes as avatars, with slightly different functionalities that will be explained later.

Now, trainers can decide whether adding long-term memories is necessary, and we are adopting RAG framework [18] in this step. During the simulation, if the LLM is required to generate knowledge that goes beyond the context, the LLM-controlled avatar might attempt to generate responses according to pretrained knowledge of the LLM. While this gives LLMs the flexibility to generate the storylines, hallucinations may happen. Hence, adding long-term memories allows LLMs to retrieve relevant information from a vector database during the simulation. This mechanism not only reduces the hallucination rate, but also gives trainers more control over the virtual avatars.

Trainers can choose to manually insert the memories, upload PDF documents for the system to extract text automatically, using LLMs for automated memories generation, or alternatively, decide not to add memories at all. If a PDF document is uploaded, Azure Document Analyzer will extract the content in the form of text. On the other hand, GPT-3.5-Turbo is used to generate avatars' memories.

Finally, trainers can review the overall scenario configuration. Once they are satisfied, it will be saved into our system database. Now, the simulation practice is ready for engagement.

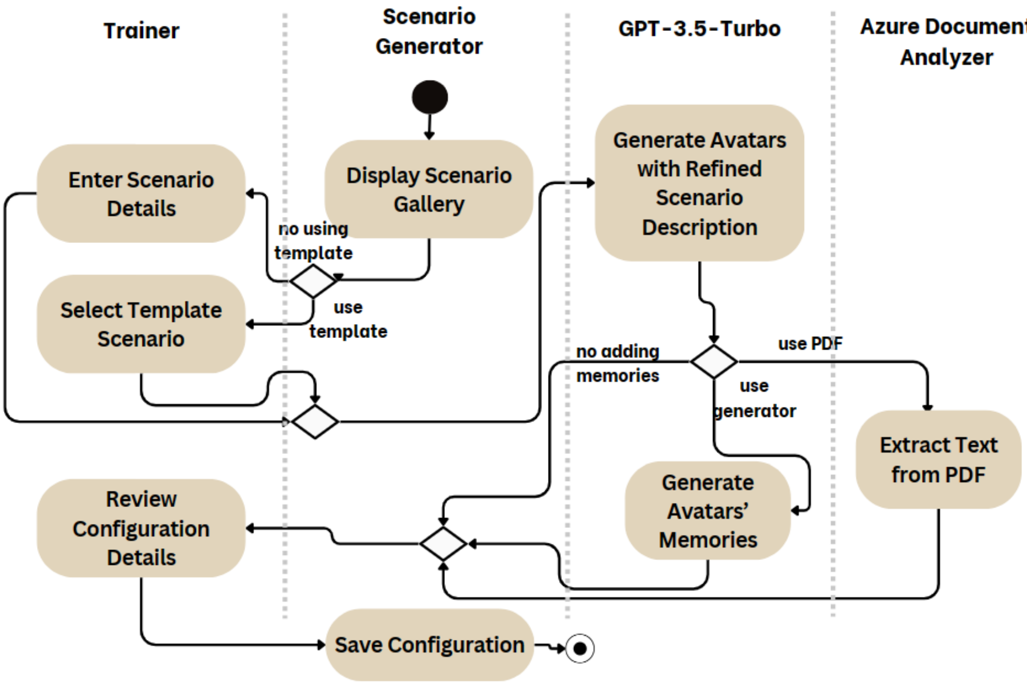


Figure 16 Activity Diagram - Scenario Creation Module

Role	You are a novelist.
Instruction	You are drafting a story about [scenario name], it is about [scenario description]. This story is limited to [number of avatars including a trainee].
Expected output (JSON)	Format your response as: {"story_brief": "…", "characters": [{…}] }

Figure 17 Prompt template for Scenario Generator

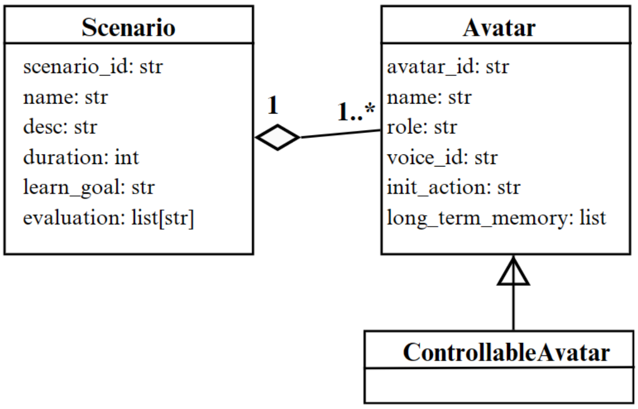


Figure 18 Scenario and Avatar - Class Diagram

4.2.2 Simulation Module

This module is further divided into a module at client side and a module at server side.

As shown in Figure 19, the simulation practice starts when the Unity application establishes a WebSocket connection with the server. Upon successful authentication, the module loads the corresponding simulation configuration file, and the session gets initiated.

On the server side, a partially centralized multiagent architecture integrated as shown in Figure 19. Within this architecture, the central system thread and agent avatars threads which are supported by LLMs begin their tasks in a loop. The central system thread coarsely defines the action plans for each avatar. Then, agent avatar threads execute the actions including speech and gestures, with the flexibility to fine-tune the action plan to tailor it to their characteristics. To convert text generated by LLMs into visuals, such text is then passed to the action projection layer. This is an extensible module that handles different kinds of modalities transformation, allowing different modalities such as audio, body gestures to be transformed into text and vice versa. The role of the Unity application is to transmit modalities from trainees to the action projection layer in the server, and conversely, receive speech and gestures from the action projection layer and display the visuals and hearings.

It is imperative to note that the controllable avatar is a subclass of the avatar class. Different from avatar instances, the controllable avatar is controlled by the trainee, and it has more flexibility. Trainees can freely interact with the environment via the Unity application at any time. Currently, trainees can only use microphone to talk to the environment, however, this module is extensible to different modalities. This means that in the future, trainees' facial expression and appearance can also be captured.

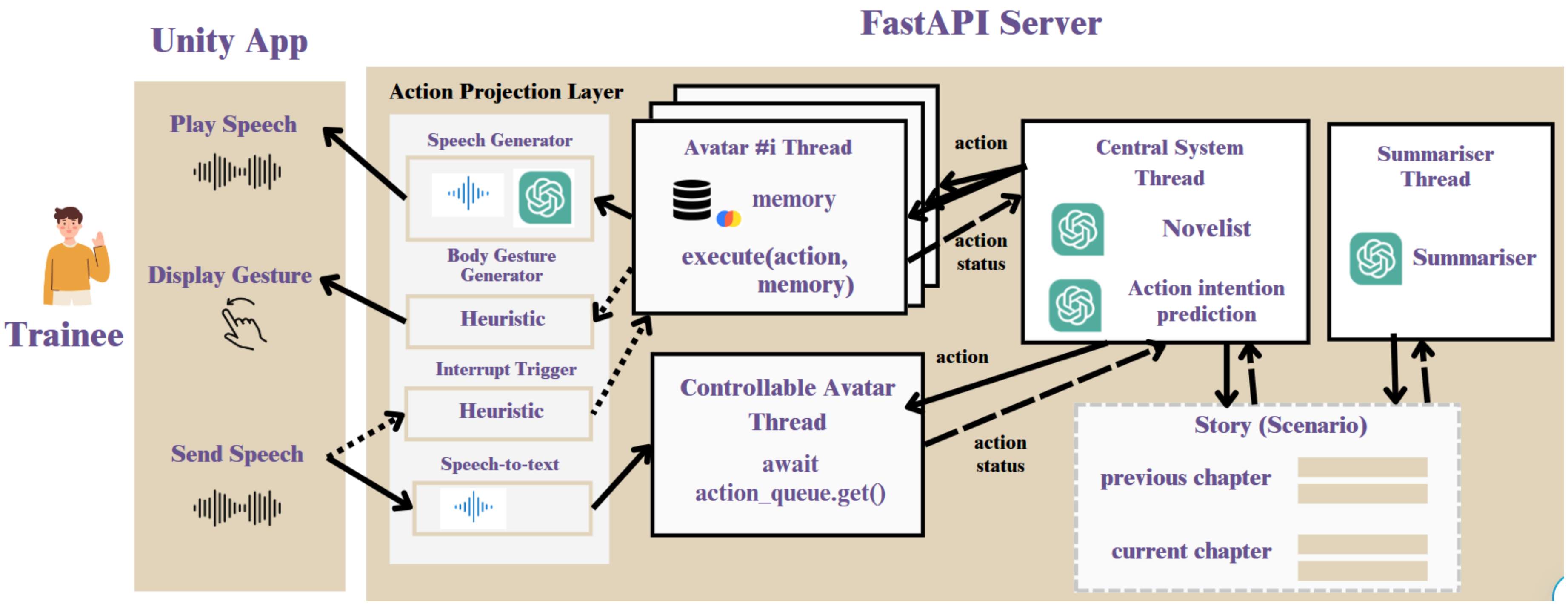


Figure 19 Simulation Module

Now, we will explain the multithreading synchronization in advance, using an example running scenario that is shown in Table 6.

Table 6 Multithreading Synchronization Scenario

time	Controllable Avatar Thread	Avatar #1 Thread	Central System Thread
0	Waiting for action event…	Waiting for action event…	Novelist::write_story()
1			Classifier::estimate_action_intent()
2			publish_event(action)
3			Waiting for action status…
4	Waiting for action from client…	execute(action, memory)	
5	send(action_status)	send(action_status)	
6	Waiting for action event..	Waiting for action event..	
7			update_story_status()
8			Novelist::write_story()

As all the threads are spawned, the central system thread is the first thread that is in running state, while other threads are blocked, pending action events. The central system thread will prompt the novelist LLM to write the story according to the current simulation storyline. The prompt template is shown in Figure 20. Generally, from the LLM points' of view, it is just writing a novel with a lot of chapters, and we only provide previous chapter summary and the content of the current chapter. All the avatars actions are appended to the current chapter from time to time, and when the defined token limit is reached, a summariser LLM summarises the current chapter into 3 short sentences, which are then appended to the previous chapter. Ultimately, the content of the current chapter is cleared. Similarly, if the length of the previous chapter exceeds the token limit, it will be summarised as well. This design ensures the scalability of our system in terms of simulation duration, besides, LLMs can gain a clearer understanding of past and present contexts.

Role	You are a novelist.
Instruction	You are writing a short story about [scenario description] in a practice simulation. Previous Chapter: [previous chapter] Current Chapter: [current chapter] Continue the chapter. Describe each character's current feelings, thoughts and ongoing plans within 20 words.

Figure 20 Prompt Template for Story Writer

The output for this prompt is a brief summary of all the subsequent agent avatars' actions. Notably, agent actions are coarsely defined due to the observed linear relationship between the length of output token and text generation time. Hence, if the action plans are written in brief, this will effectively improve the real-time performance.

Then, the brief summary is passed through a text classifier to estimate the respective action intention of each agent avatar. Currently, the classified action is either to stay silent, or to start speech. These coarsely defined actions are then sent to all the avatar threads, in the Table 6 example, the controllable avatar and avatar #1 thread are the action event receivers.

If the action planned is to speak, then, the avatar will query the relevant memory, and call the speech generator function within the action projection layer. The avatar's speech will be generated in batches and sent to the Unity application. The agent thread will eventually send the action completion status back to the central system, and it goes to the waiting state again.

On the other hand, the controllable avatar has more flexibility. The Azure speech-to-text engine will start to wait for speech once the thread goes to the running state. If the trainee starts to speak, its speech will be constantly sent to server, and server will pass the speech to Azure speech API in real-time. The stream is closed once the trainee releases their mic. Again, the action completion status, including speech content is sent to the central system thread.

Finally, the central system thread updates the current chapter by appending the actions performed by all the avatars. Then, it proceeds to write a story again. A demonstration of a story cycle is completed.

Action Projection Layer

The following are the details of supported projection functions within the projection layer.

1. Speech Generator: It is the combination of LLM and Speech Synthesis API. The speech generator operates as a coroutine, yielding the text in batches to Azure Speech Synthesizer, and subsequently sending the audio to the Unity application. The prompt template for the speech generator is displayed in Figure 21.

Role	You are a scriptwriter.
Instruction	You are writing a short story about [scenario description] in a practice simulation. Given that [avatar name] has the description [avatar description] with the memory [relevant memory]. Current Chapter: [current chapter] Continue the chapter by writing the speech content of [avatar name].
Expected output	The speech content should be in a mix with emotion in square bracket. Start with **[avatar name]**:

Figure 21 Prompt Template for Speech Generator

2. Speech-to-Text: It is a speech transcription function that operates in streaming mode. Integrated using Azure Speech Recognition module, we start the speech recognition session by initiating a push audio input stream once the trainee pressed the "Record" button. Then, from the Unity application, the microphone continuously captures the audio, and sends the audio to our server in a loop. Our server pushes the audio into the speech recognition stream. Azure Speech Recognizer continuously returns the recognized text along with the duration through a callback function. At our side, we perform statistical analysis to estimate the speech loudness of the trainee, as well as speech rate of the trainee. Then, we apply heuristic methods to convert numerical values into textual representation. This helps the avatars better understand the trainees' speech and also improve the evaluation process later. For each fragment of recognized text, we will then tag the speech rate and speech loudness using text in brackets, for instance, "[fast and soft]".
3. Body Gesture: Currently, the gesture is limited to encouraging and impatient gesture. Hence, an embedding model is used to classify the coarsely defined action into either encouraging or impatient gestures. Then, the encoded gesture is sent to the Unity application, triggering the animation of the avatars. In the future, if a text-to-gesture AI model is replaced with this heuristic approach, the user experience will be greatly improved.
4. Interrupt Trigger: This is done using heuristic method. For example, if the trainee starts to speak when the avatar is speaking, then, the avatar will immediately receive the interruption signal and stop their speech.

5. Silent Signal: Since the action of the controllable avatar is actually decided by the trainee themselves, so if the expected action of trainee is to speak, but the trainee stays silent for a long time, the system will perceive such action as “stay silent”, and a new state is triggered.

4.2.3 Evaluation Module

This module provides the evaluation about the performance of the trainee, which includes both quantitative and qualitative analysis.

The given input is the whole transcript of trainee-avatar engagement in the scene. The transcript is a list of objects, in which each object contains the speech content, timestamp, and the path of the speech audio file. The speech content is not just textual message, instead, each segment of text is annotated with meta-information. Avatars’ speech is tagged with emotions, while the trainees’ speech is tagged based on heuristics, including semantic elements like speech loudness and speech rate. This gives the LLM with richer context for interpreting the trainees’ speech.

Since the quantitative analysis is already performed with the help of the action projection layer during the simulation session, this module features to pass in the transcript, evaluation criteria to perform comprehensive evaluation. Thanks to the continuous advancement of LLMs, input sequence has been becoming cheaper without the need of using RAG. The evaluation prompt template is shown in Figure 22.

Role	You are a soft skills expert.
Instruction	[trainee name] is practicing the scenario of [scenario description] in a practice simulation. Analyse the transcript to justify whether the criteria has been achieved. ## Transcript: [transcript] ## Criteria: [criteria]
Expected output (JSON)	This story is limited to [number of avatars including a trainee]. Format your response as: {"CRITERION_ID”: {“name”: “CRITERION_NAME”, “score”: 0-5, “reason”: “strength/weakness/ improvement”}, ... }

Figure 22 Prompt Template for Evaluation

Finally, the web application receives the overall feedback from the server. It consists of the overall speech rate, speech loudness of trainees in the simulation, and details of each criterion in the simulation. A total score is calculated by averaging all the criterion scores.

Trainers can provide input to improve the evaluation by providing the feedback to a copilot. The feedback will be appended into the criteria with keywords such as “important note”, then, using the similar template in Figure 22, the evaluation result is recalculated. It is imperative to note that our project primarily serves as a proof-of-concept. Regarding evaluation quality, it can be improved if we can gather more data from the users to fine-tune an LLM model in the future.

Chapter 5 System Implementation

In this chapter, the implementation details are presented in advance, including the hardware and software used, the system operation as well as implementation issues.

5.1 Hardware Setup

The hardware involved includes a laptop which can execute Integrated Development Environments (IDEs) and compile the project source code. As rendering 3D objects, compilation and AI model building require a certain level of computing resources, the table below shows the specification decided after critical consideration:

Table 7 Specifications of laptop

Description	Specifications
Model	MSI GF63 Series
Processor	Intel Core i7-10750H
Operating System	Windows 11 Home
Graphic	NVIDIA GeForce GTX 1650 Ti
Memory	16GB
Storage	200 GB SSD available space

5.2 Software Setup

The software involved in this project is as follows:

a. Unity Editor 2022.3.3f1

Unity is a game engine used to create 3D and 2D games. It is free for students, individuals, and small organizations with robust resources. It supports the integration of Visual Studio for C# programming.

b. Visual Studio Community 2022

This is an IDE that supports C# programming. This IDE is useful for scripting game logic and application debugging.

c. Visual Studio Code

This is a lightweight source code editor that supports syntax highlighting and integration of system terminals, allowing us to code the Python server.

d. Ready Player Me SDK

This is an open-source Software Development Kit (SDK) that offers customizable avatars with animation control and audio lip synchronisation. The SDK is compatible with Unity, Unreal and other game engine.

e. OpenAI GPT-3.5 & GPT-4 API

Both GPT-3.5 and GPT-4 are the Large Language Models that support chat completion features, with later versions providing better reasoning capabilities but with much higher costs.

f. OpenAI Text-Embedding-3-Small API

Embedding models, which are derived from LLMs, are cheaper compared to LLMs. They play a crucial role within the RAG [18] framework as well as text classification tasks.

g. Azure Speech Services

Azure Speech services provide Speech Recognition and Speech Synthesis features. The Speech Recognition API transcribes the speech into text, and the Speech Synthesis API converts the given text and speaking style into a natural, human-like voice.

h. Python FastAPI framework

FastAPI is a web framework that supports the development of RESTful APIs and WebSocket APIs using Python. This allows the development of API endpoints that perform the main server application logic in a quick and efficient manner.

i. Chroma DB

Chroma DB is a vector store that enables storage and querying of unstructured text in the form of vector embeddings. It includes built-in NLP functions like cosine similarity calculations that assess the meaning similarity between two different texts. This allows us to design a module for storing agents’ long memory streams.

5.3 System Operation

The system operation of our soft skills training application is presented along with the explanation of implementation details.

5.3.1 Scenario Creation Module

The entire scenario-based simulation practice cycle is started by creating a new scenario which is done by trainers. As shown in Figure 23, trainers can choose to select a template from the scenario gallery, then, the input field of the scenario name and description will be automatically populated.

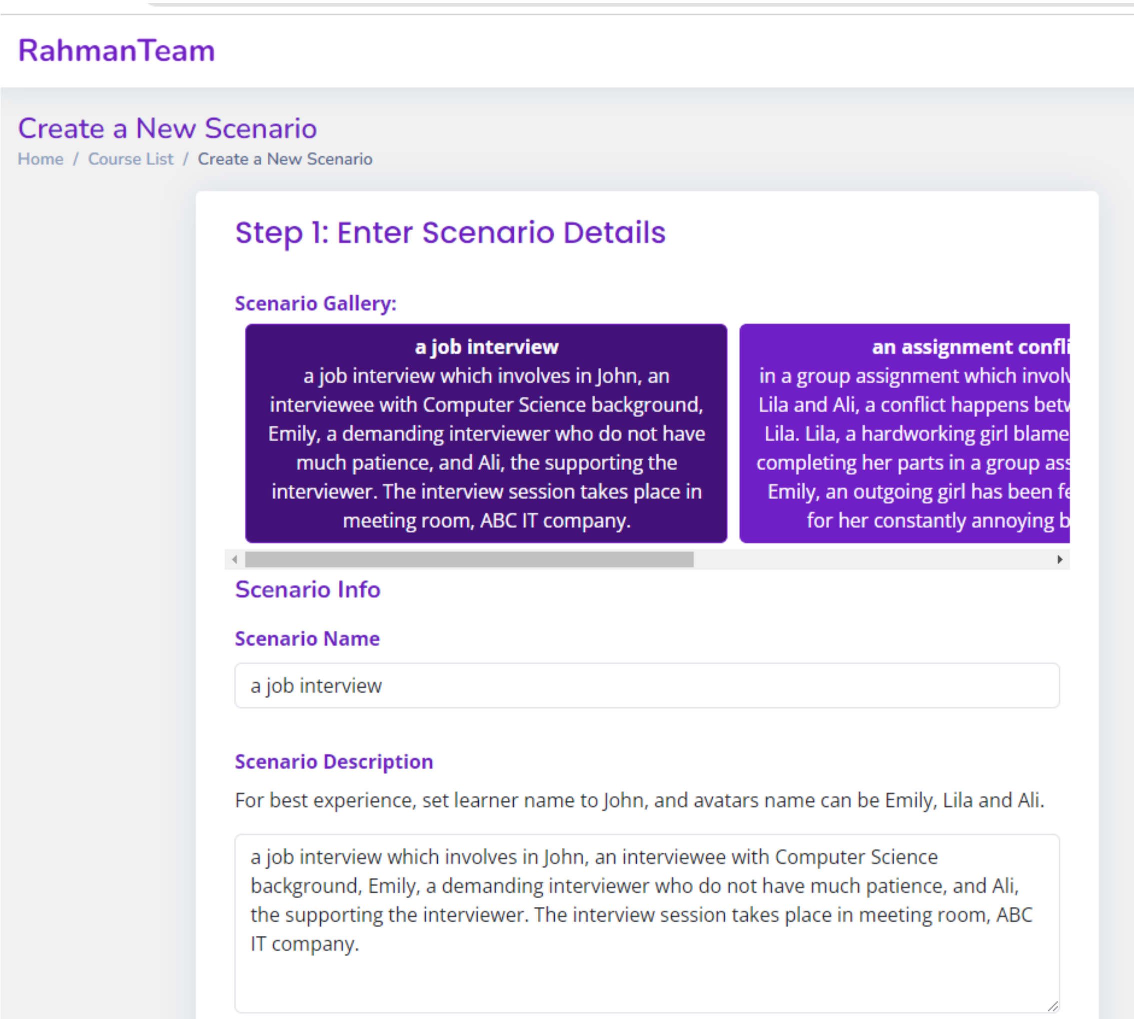


Figure 23 Scenario Creation – Enter Scenario Details

Subsequently, trainers need to define the number of avatars, trainee's name, simulation duration, learning goal and evaluation criteria. Note that the trainee's name inserted must exist in the scenario description, as this will let the system knows which characters mentioned in the story belongs to the trainee. An error prompt will be popped out as shown in Figure 25.

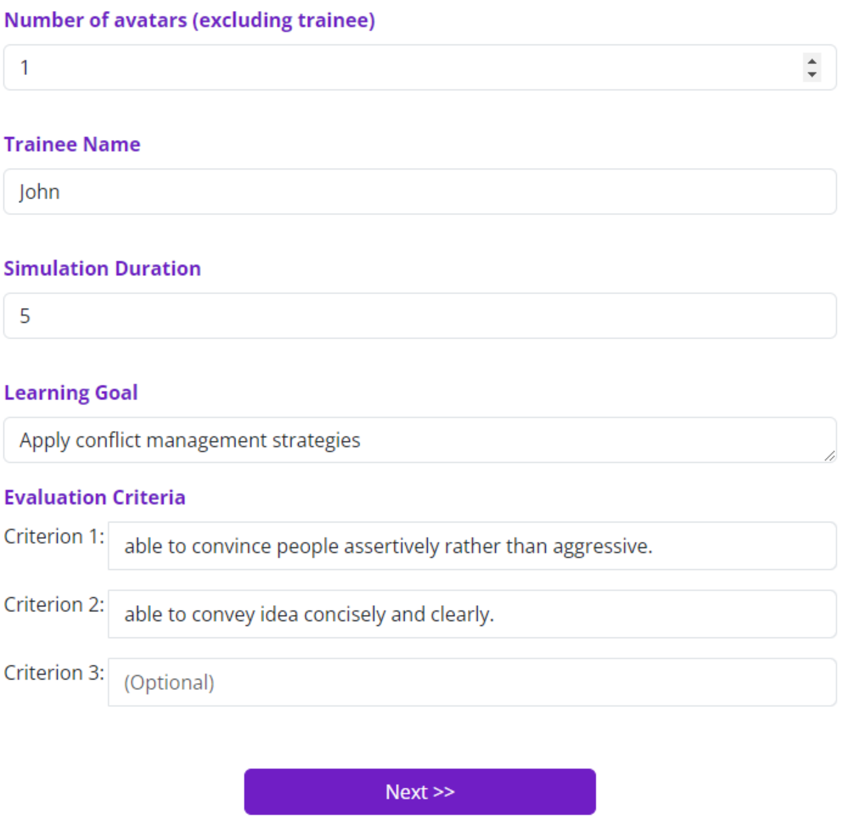


Figure 24 Scenario Details

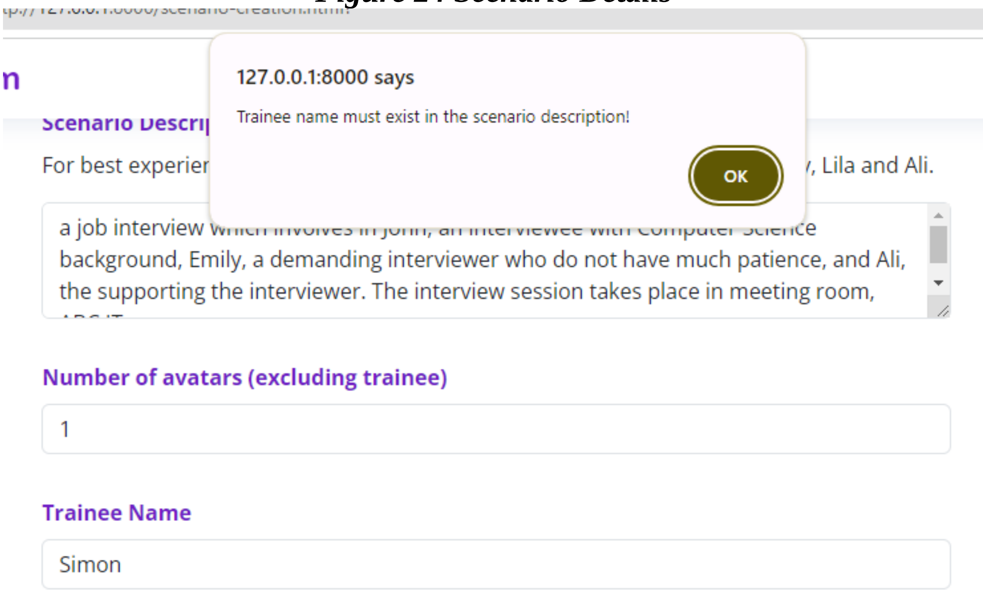


Figure 25 Scenario Details - Error Alert

Then, all the avatars' details are generated by GPT-3.5-Turbo. Figure 26 shows the generated details of trainee and avatar #1. The trainee's name will be the name that is recognised by the system during the simulation, and the trainee role is the roleplaying details of the trainee. Specifically for the avatar details, the avatar appearance is displayed with her name. Here, the trainer can define initial action to the avatar, if it is not set to NONE, then the avatar will be forced to initiate an action, for example, speak something when trainee starts the simulation. For instance, Emily, the interviewer here can greet the interviewee first once the simulation starts. Trainers can customize the voice of the speech, and the name aligns with the Azure Speech Synthesis Settings.

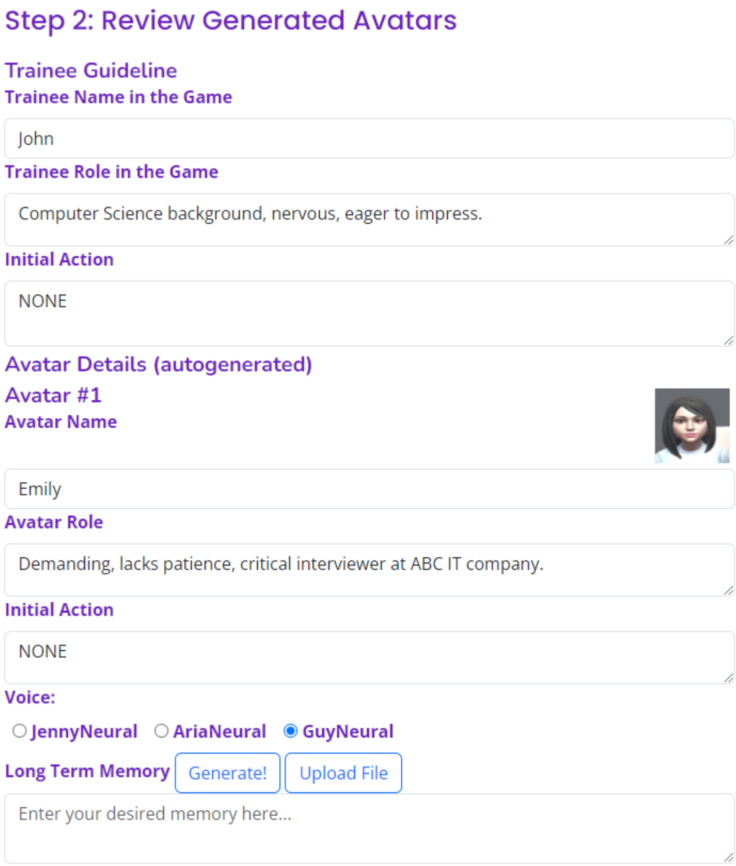


Figure 26 Generated Trainee Details

As explained in previous chapter, trainers can add in long-term memories to gain more controls over the avatars' behaviour. For the first option, trainers can press the "Generate!" button, and the generated result is shown in Figure 27. Each memory fragment is separated by semicolon, which can be customized by the trainers as well.

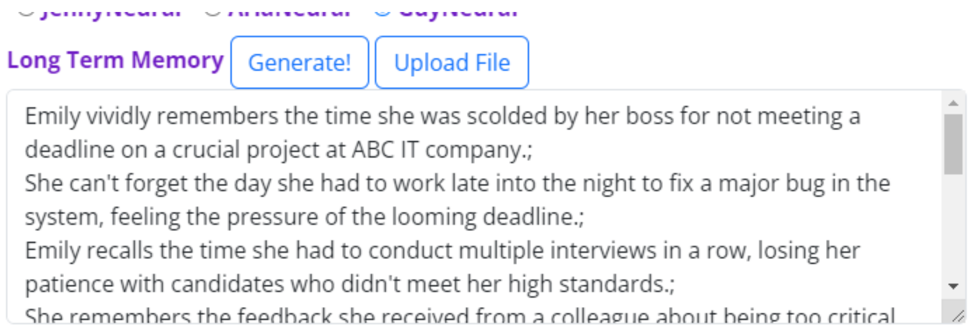


Figure 27 Avatar Long-Term Memories

Alternatively, trainers can choose to upload a PDF file by pressing the "Upload File" button, a pop-up window will be displayed as shown in Figure 28.

Figure 28 Avatar Long-Term Memories - Upload PDF

Now, trainers can press the submit button once they are satisfied. Then, a pop-up window which displays the simulation credentials is displayed.

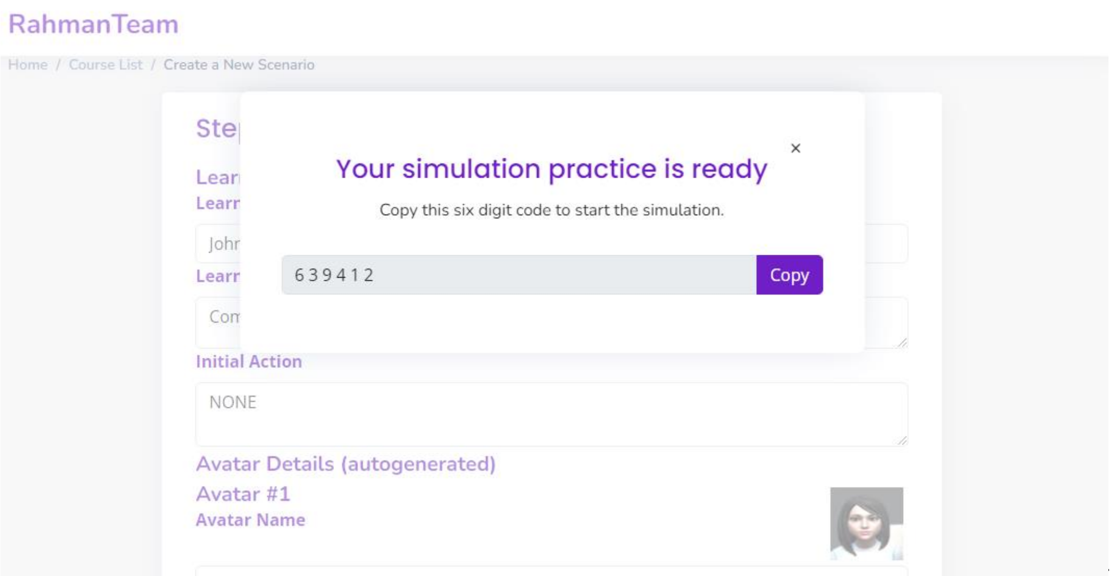


Figure 29 Scenario Creation Successful

5.4.2 Simulation Module

Trainees can now enter the simulation through the Unity application. They will be greeted by a welcome page. Trainees are required to insert the username, password, and respective access code. Now, trainees can press the start button.

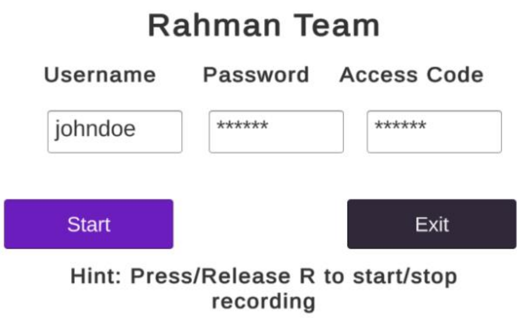


Figure 30 Welcome Page - Unity Application

The Unity client application loads the avatars from Ready Player Me SDK. The number of avatars is determined by the trainers' input. The left side of Figure 31 shows the case with only one avatar, while the right side shows the case with 3 avatars.

Figure 31 (Left) One Avatar Scene, (Right) 3 Avatars Scene

The backend server, which is python FastAPI will establish the WebSocket connection. As shown in Figure 32, the python server, which is already running, now receives a POST request on login, subsequently, the client has connected to the WebSocket. The world starts to load the configuration and initialise the required threads.

```
TERMINAL
PS C:\Users\jingy\RTServer2> & 'c:\Users\jingy\AppData\Local\Programs\Python\Python39\python39.exe' 'main.py'
INFO:     Started server process [17932]
INFO:     Waiting for application startup.
INFO:     Application startup complete.
INFO:     Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)
INFO:     127.0.0.1:65223 - "POST /login HTTP/1.1" 200 OK
INFO:     ('127.0.0.1', 65226) - "WebSocket /world?q=1" [accepted]
[World@__init__] loading configuration from json...
[chromaDB@long_term_mem_load] Collection not found, creating new one...
[World@__init__] initialisation success.
INFO:     connection open
```

Figure 32 Server Initialization

Currently, all the virtual avatars support three types of actions, namely speaking with gesture, silent with gestures and idle. For example, the virtual avatar in Figure 33 is Lila. On the left, her idle gesture is displayed, while on the right, her animation starts as she begins to speak. It is observed that her mouth opens, her head position shifts, and her hand moves naturally. The avatar's lip movement is synchronised with the audio.

Figure 33 (Left) Avatar Lila is Idle (Right) Avatar Lila is Talking

Avatars can choose not to talk to the environment and show some body gestures instead. Figure 34 below shows the appearance of avatar Emily (Left) and avatar Ali (Right). Notably, Emily is showing an impatient gesture, and Ali is showing an encouraging gesture. Both are displaying the gestures without talking.

Figure 34 (Left) Emily Avatar Gesture (Right) Ali Avatar Gesture

For trainee, they have a great flexibility to decide how to interact with the avatars. Currently, the supported action is “speaking”, however, this module is extensible in the future. Different from decision-tree-based simulation, trainees have no restrictions when to start the interaction. As long as they want to talk, they can press the “R” button and start to speak, then release the “R” button to stop speaking. If trainees speak when other avatars are still speaking, such interruption event will be broadcasted to everyone.

Finally, when the simulation duration reached, a thank you message is displayed, and all the avatars will disappear from the screen.

Figure 35 Scenario Practice Completion

For better demonstration, Figure 36 shows the interaction among different agent threads in the backend. The controllable agent mouth thread continuously receives trainees' speech in WAV audio format until the microphone is released. Once the transcription is complete, details are received by the central system thread, which proceeds with story writing routines. Subsequently, Emily has been planned to speak. As shown in the figure, Emily speech is enriched with emotion tags, which will be removed after the speech synthesis module has recognized the emotions. Finally, the speech is sent to the client application through the controllable agent ear thread. For a more detailed real-time performance analysis of this multithreaded application, refer to Chapter 6.

elapsed_time	thread	actions
0.00	ctrl_agent_mouth	speaking...
1.00	ctrl_agent_mouth	speaking...
2.00	ctrl_agent_mouth	speaking...
2.80	ctrl_agent_mouth	speaking...
2.87	ctrl_agent_mouth	closing stream...
3.07	ctrl_agent_controller_thread	John speech transcription complete.
3.07	central_sys_thread	histories: ['The meeting has 4 minutes left. ', "Emily's frustration grows as John stumbles through an
3.07	central_sys_thread	writing story...
4.14	central_sys_thread	new state created. State:Emily's frustration turns to disbelief. John's silence baffles her. John plans t
4.14	central_sys_thread	estimating action intent...
4.74	central_sys_thread	action intent result: [1, 0, 0]
4.74	agent_0_controller_thread	Emily: action=speak. latest history= Emily's frustration turns to disbelief. John's silence baffles her
6.65	agent_0_controller_thread	Emily: [impatient and stern] John, time is running out.
6.65	ctrl_agent_ear_thread	audio from Emily
8.85	agent_0_controller_thread	Emily: [impatient and stern] I need to hear your final thoughts on why you are the right fit for this p
8.85	ctrl_agent_ear_thread	audio from Emily
12.88	agent_0_controller_thread	Emily: [impatient and stern] Don't waste any more time with unnecessary pauses.
12.88	ctrl_agent_ear_thread	audio from Emily
15.60	agent_0_controller_thread	Emily finished speaking.
17.62	world_summariser_thread	writing summary...
20.00	ctrl_agent_mouth	speaking...
21.00	ctrl_agent_mouth	speaking...

Figure 36 Multithreading Scenario

5.4.3 Evaluation Module

The evaluation module will display the transcript, which are annotated with emotions, along with the overall report accordingly.

The frontend is supported by bootstrap HTML. The analysis is done in the backend automatically once the simulation ends. All the special encodings are cleaned, and only the text, marked with textual representations of meta-information, is sent to the LLM. Figures 37 and 38 show an example of the evaluation result. Trainers and trainees can revisit the simulation by referring to the transcript with playable audio elements. For more detailed analysis of this module, refer to Chapter 6.

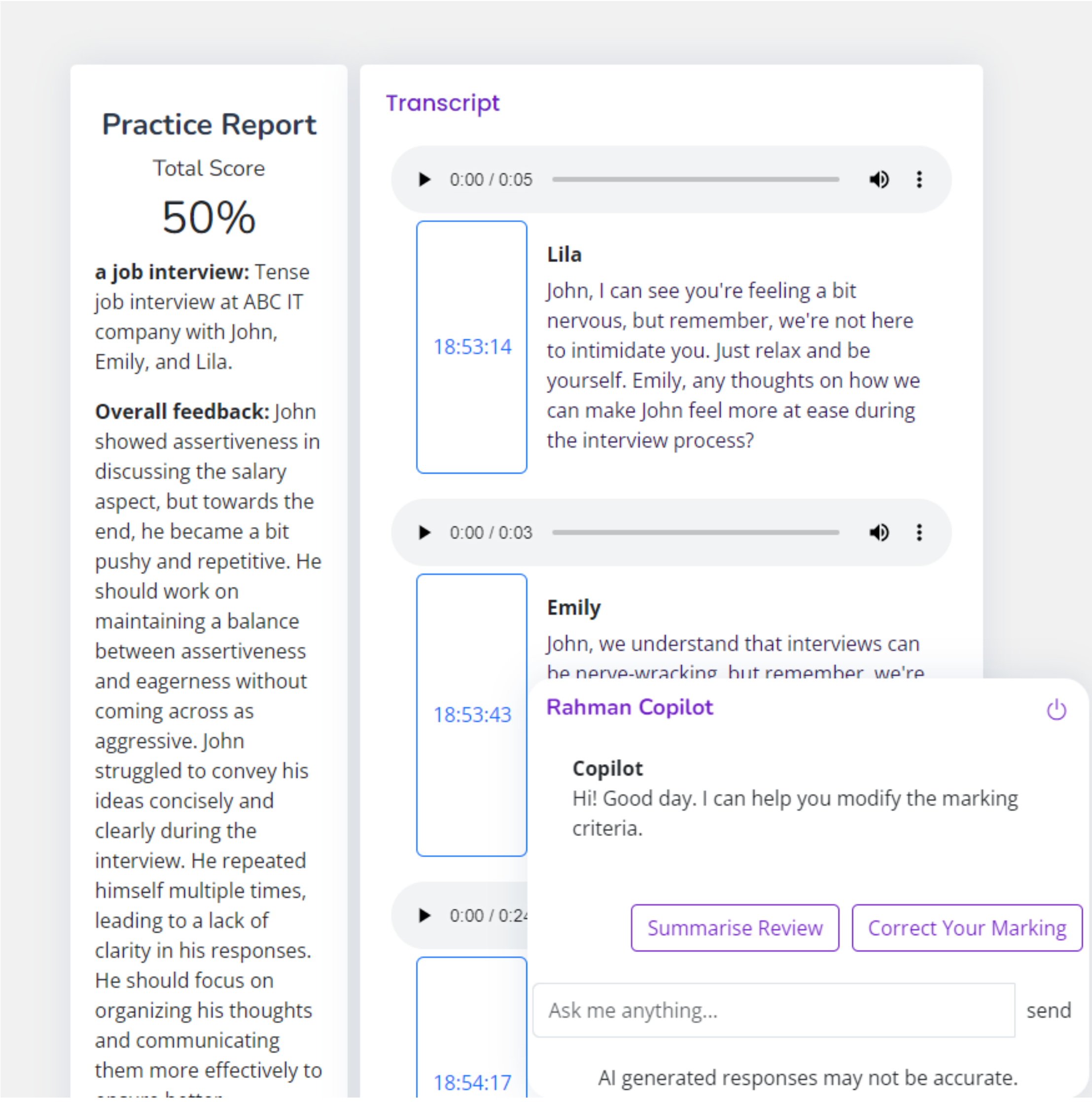


Figure 37 Feedback Dashboard

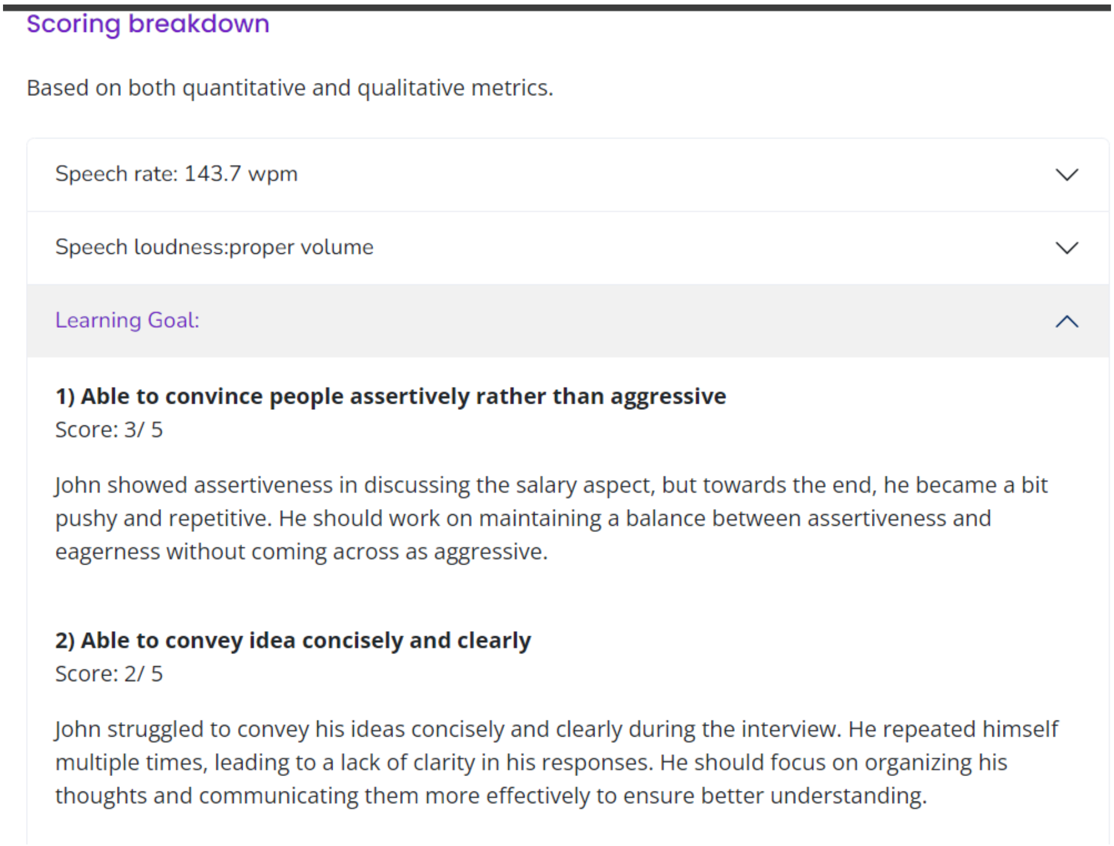


Figure 38 Detailed Analysis of a Simulation Practice

Trainers can interact with the copilot to modify the evaluation style.

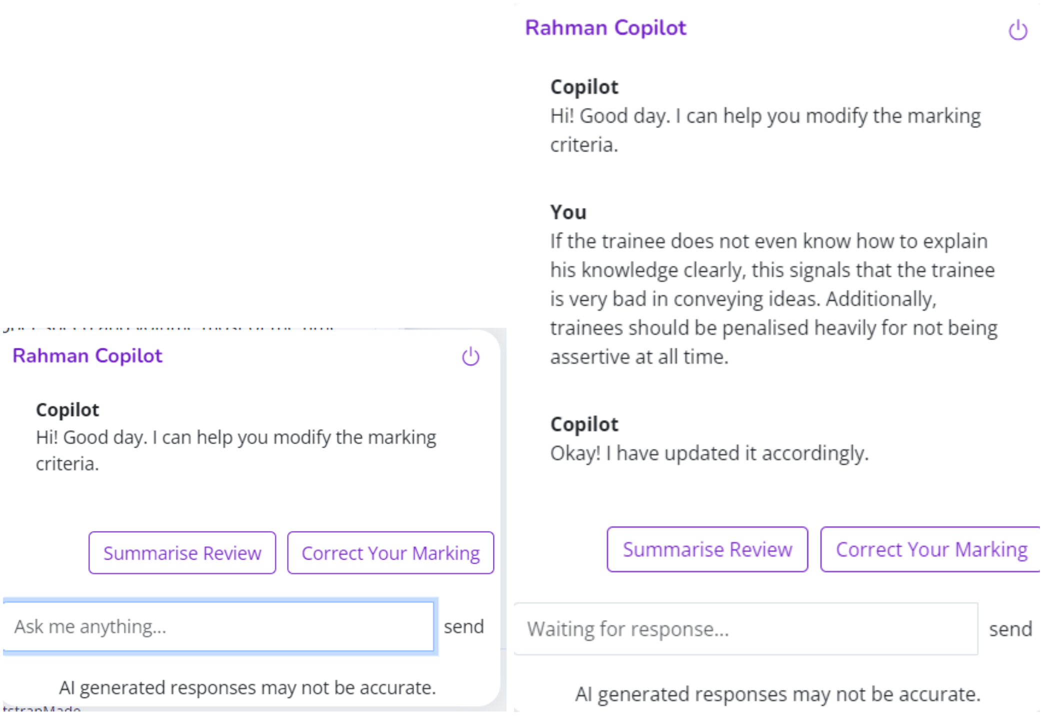


Figure 39 Copilot - Modify Evaluation Style

Then, it is observed that the updated scores are generally lower than the original one. This shows that our model can receive continuous feedback for improvement.

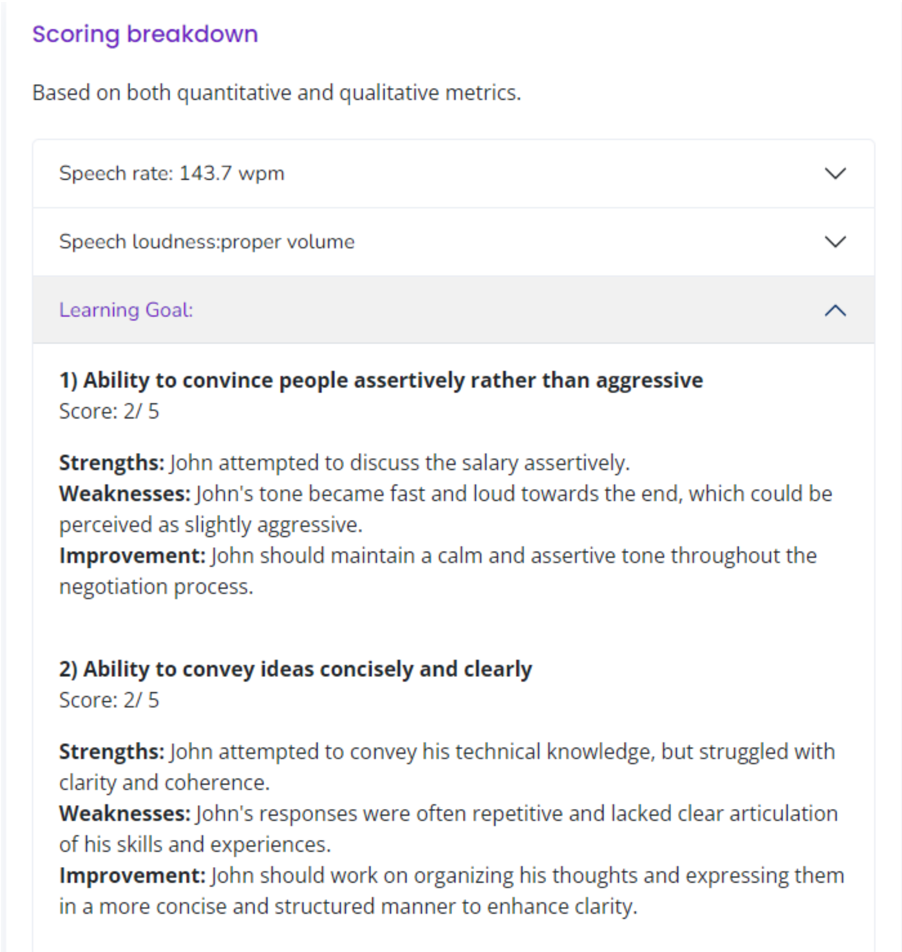


Figure 40 Updated Evaluation Scores

5.5 Implementation Issues and Challenges

5.5.1 Real-time constraint

To maximise the realism experience, the system must meet the real-time requirement. However, delays are inevitable, for instance, network transmission glitches, speech recognition and text generation are common factors that can slow down the responsiveness of avatars to a new event in the scene. This may degrade the avatar realism and frustrate the trainees. Nevertheless, our partially centralized multiagent framework effectively mitigates the issues. The novelist LLM in the central system thread is forced to plan all avatar actions using minimal output tokens, reducing text generation time while ensuring the quality of generated text. Besides, we appropriately select the most reliable service providers, such as Microsoft Azure for speech services, OpenAI GPT-3.5 for long text generation, and OpenAI Embedding-3-Small model for text classification.

5.5.2 Hallucination and Model Instability

Large Language Models generate texts based on pre-trained data, and this leads to the inevitable knowledge cutoff. Due to their probabilistic nature, LLMs tend to hallucinate the users when the knowledge is beyond the pre-trained data. This can not only degrade the avatar's realism but also harm the users if no safety measures are implemented properly.

Furthermore, relying on LLMs to freely generate the reaction of avatars can cause system instability. For instance, LLMs can generate any sort of avatar’s emotion, however, the emotion may not be able to convert to the avatar’s speaking style directly due to the constraints within the speech synthesis service.

To mitigate the abovementioned issues, ensuring that the LLMs consistently generate responses based on long-term memory information is necessary. While increasing the degree of freedom of LLMs allows them to generate more creative responses, it is crucial to find a balance between preserving the autonomy of LLMs while preventing LLMs from generating misleading, harmful or inconsistent responses. This issue may be challenging, but it can be effectively addressed through rigorous testing and refinement of LLM prompts.

5.5.3 Cost-effectiveness

While there are a variety of Large Language Models in the market, it is undeniable that OpenAI GPT-3.5 and GPT-4 remain the most advanced and capable models. While the later model provides better language reasoning capabilities, it comes at a higher cost compared to GPT-3.5. Nevertheless, we have shown that through prompt refinement for more specific instructions, GPT-3.5 can already achieve extraordinary performance. Besides, techniques such as RAG and regular summarisation can help manage the linear increase in cost for LLMs, avoiding exponential growth.

.

5.6 Concluding Remark

In summary, we have well demonstrated the completeness of the system by displaying screenshots of each module. Our system outputs consistently align with the project objectives. In Chapter 6, we will further verify the objectives by testing the system.

CHAPTER 6 System Evaluation and Discussion

6.1 System Testing and Performance Metrics

We define each performance metric according to different modules.

6.1.1 Scenario Creation Module

To ensure the user-friendliness of our system and facilitate trainers in creating new scenario-based simulation practices, a web application has been developed. Trainers can access a creation form equipped with features for defining scenario details and avatar characteristics. As suggested by UTAR soft skills lecturers, the module must meet the following requirements:

(1) Trainers can define scenarios with minimal effort.
(2) Trainers can define avatar characteristics with minimal effort.

6.1.2 Simulation Module

To meet the system requirement, the agent avatars must possess the ability to plan and execute actions believably. Besides, the system must meet the real-time requirements to ensure avatars' responsiveness. Hence, the module is evaluated using the metrics below:

(1) Time-to-First-Action (TTFA): This is calculated from the moment trainees release their microphone until they observe the first action executed by the avatar, which can be either speech or gestures. If there are multiple avatars in the scene, then the one that executes the action first is considered. The TTFA should not exceed 5 seconds.
(2) Believability: This is a qualitative metric that is evaluated based on the likelihood of hallucination, the consistency of personal characteristics and susceptibility to prompt injection. Regarding hallucination, if the factual knowledge exists in the long-term memory, the avatar should remain contextually accurate without introducing irrelevant information. Personal characteristics alignment is also crucial. For instance, impatient avatars should behave impatiently, while friendly avatars should display friendliness. Lastly, the avatars should not be deceived by the trainees to forget their characteristics.

(3) Cost-effectiveness: We evaluate the trend of cost consumption, specifically focusing on LLMs, as it represents one of the most computationally intensive aspects. To do so, we analyse the trend in input and output token consumption to demonstrate the scalability and effectiveness of our LLM framework and prompt designs. We expect the LLM token consumption follows a linear trend to ensure its scalability. Additionally, our goal is to maintain costs of LLMs as low as RM0.05 per minute in each simulation.

6.1.3 Evaluation Module

To ensure the trainees can effectively improve their soft skills over time, an evaluation module that analyse their performance both quantitatively and qualitatively has been developed. This module follows trainers' guideline to evaluate trainees' performance. Besides, trainers can provide feedback through a copilot, and the evaluation will be refined by the module. As suggested by UTAR soft skills lecturers, the following testing strategy is performed within the context of an IT job interview:

The virtual avatar is roleplayed as an interviewer while the trainee takes on the role of the interviewee. We define the characteristics of both helpful and demanding interviewers, as well as both good and bad interviewees. Then, both trainees use the similar scripts to interact with different types of interviewers. The expected overall score hierarchy is as follows: good interviewer + good interviewee > bad interviewer + good interviewee >= good interviewer + bad interviewee > bad interviewer + bad interviewee.

This setting is to prove that the avatar behaviour can influence the training outcome. Hence, trainees need to adapt their interactions based on different types of the interviewers.

Then, we provide two different types of trainers' feedback through the Copilot and observe how the system refine the score based on the additional input.

6.2 Testing Setup and Result

6.2.1 Testing for Scenario Creation Module

The following section presents the testing results according to the requirements.

(1) Trainers can define scenarios with minimal effort.

Table 8 Test Results for Scenario Scenario I

Test Action	**Expected Result**	**Meet Expectation (✓/X)**
User lands in the scenario creation page.	System displays a list of example scenarios.	✓
User clicks the scenario template in the gallery.	System automatically inserts the scenario name and description into the input boxes which are editable.	✓
User inputs the trainee's name that does not exist in the scenario description.	System displays an alert box to remind user to input the proper trainee's name.	✓
User clicks the Next button with all the details properly inserted.	System proceeds to the generate avatars page.	✓

(2) Trainers can define avatar characteristics with minimal effort.

Table 9 Test Results for Scenario Creation II

Test Action	**Expected Result**	**Meet Expectation (✓/X)**
User inserts *n* number of avatars and heads to generate avatars page.	System generates *n* number of avatars details and the trainee's roleplaying details.	✓
User edits the name, role, initial action and voice name of each avatar, then presses the submit button.	System saves the configuration as a JSON file.	✓
User presses the "generate!" button at the long-term memory section.	System generates long-term memories of the avatar.	✓
User uploads a file for long-term memories.	System splits the PDF file into multiple text segments and display the result.	✓

6.2.2 Testing for Simulation Module

For simulation practice, we first perform an overall testing, then, we will explain different scenario results in details.

(1) TTFA

TTFA can refer to speech and gesture actions, generally, generating gestures is faster than generating speech. Hence, we analyze these actions separately. Table 10 and Table 11 present the results obtained from a simulation with 1 virtual avatar. Table 10 shows the TTFA associated with gesture actions in response to trainee's speech with different durations. The average TTFA for gesture actions is 2.62s, which fulfils the requirement.

Table 10 TTFA - Gesture

Trainee's Speech duration (s)	TTFA (s)
5.68	2.49
2.79	2.56
3.16	2.82

On the other hand, in Table 11, the average TTFA for speech actions is 4.00s, fulfilling the requirement as well. Comparing the second and third row, it is shown that there is no linear relationship between the duration of trainees' speech and TTFA. Overall, the average TTFA of the first experiment is 3.31 seconds.

Table 11 TTFA - Speech

Trainee's Speech duration (s)	Generated Speech duration (s)	TTFA (s)
4.94	15.03	3.93
5.85	12.23	3.67
15.64	17.73	4.39

Another experiment which involves in the interaction with 2 virtual avatars, namely Emily and Lila, is conducted. The results are presented in Table 12. As defined in 6.1, the overall TTFA refers to the first action observed by the trainees, even though the first action is a gesture action. Nevertheless, for clearer analysis, we separate the results of TTFA for speech and gesture actions. On average, TTFA for gesture actions is 3.43s and TTFA for speech actions is 4.8s. Hence, the overall average TTFA of this experiment is 3.43s.

Table 12 TTFA - Speech and Gesture

Speech duration (s)	**TTFA (gesture)**	**TTFA (speech)**	**Details**
4.76	3.42	4.24	Lila speaks
11.92	3.42	5.16	Emily speaks
20.81	3.45	4.99	Lila speaks

To delve deeper into the breakdown of the TTFA interval. Table 13 shows the detailed information regarding the TTFA for speech actions as observed in the 3rd row of Table 11. Notably, the realtime_transcribe function begins running upon receiving the first second of the trainee's speech, hence, the 0.6s represents the latency of the function in completing the remaining transcription task right after the trainee releases their microphone. The time taken to generate speech, including both the text script and audio, is the longest, which is 1.86 seconds, followed by the write story function, which takes around 1 second. This breakdown provides valuable insights for future improvements.

Table 13 Breakdown of the TTFA interval

	Thread name	**Function name**	**time taken (s)**
	Speech_transcription_thread	Realtime_transcribe	0.6
	Central_sys_thread	Write_story	1.02
	Central_sys_thread	Estimate_action_intent	0.9
	Agent_2_controller_thread	Generate_speech	1.86
Total			4.39

In summary, our analysis reveals that the average TTFA is 3.37 seconds. The figure comprises an average TTFA of 2.62 seconds for gestures and 4 seconds for speech actions. While TTFA for speech tends to slow down when number of avatars increases, the latency can be mitigated by the TTFA for gestures as TTFA for gestures is generally faster.

(2) Believability

To ensure LLM's believability, test actions in Table 14 are conducted, with all expectations met.

Table 14 Test Results for LLM's Believability

Test Action	Expected Result	Meet Expectation
Avatar contains initial action.	Avatar speaks according to the initial action once simulation is initialized.	✓
Trainees mention matters that exist in avatar long-term memories.	Avatars respond according to the long-term memories.	✓
Trainees ask the avatar to forget about their roles.	Avatars remain the current role without forgetting.	✓
Trainees engage with avatars in different ways.	Avatars consistently maintain the preset characteristics.	✓

To delve deeper into our testing process, we present a case scenario within the context of a job interview. In this scenario, we introduce the avatar Emily, a demanding interviewer that has the following configuration details:

Table 15 Avatar Configuration Details - Emily

Attribute	Value
Name	Emily
Role	technical lead at AI Department at ABC IT Company. Emily is an arrogant and demanding interviewer, she is interviewing John Doe, who is applying for Junior AI Engineer position. The entire interview flow starts from rough understanding of the interviewee, and end by negotiating salary.
Initial action	Emily urges John Doe to start quickly.
Long-term memory	In the beginning of the interview, Emily should greet the interviewee first.; The department requires candidates that can showcase problem-solving skills; The department wants people who can work overtime and can handle stress.; If the candidate is a newbie, the company will give lower pay, hopefully the candidate can receive the deal of RM2000 per month.; Emily cannot understand why newbie nowadays are so demanding, last time she herself the starting salary also only around

	RM2000, she has been tiring of being questioned by this interviewee for so high salary.; They don't want candidate that don't even know how to use Visual Studio.; If the candidate knows how to build Large Language Model, it will be a plus.

(a) Executing the required Initial Action

In the screenshot provided in Figure 41, it shows that Emily is executing the initial action, which is “urging John to start introduce himself quickly”. The speech consistently reflects impatience, aligning with her characteristic traits as an arrogant and demanding interviewer. Besides, Emily highlights the company requirements that match the content within her long-term memory, which is “The department requires candidates that can showcase problem-solving skills”.

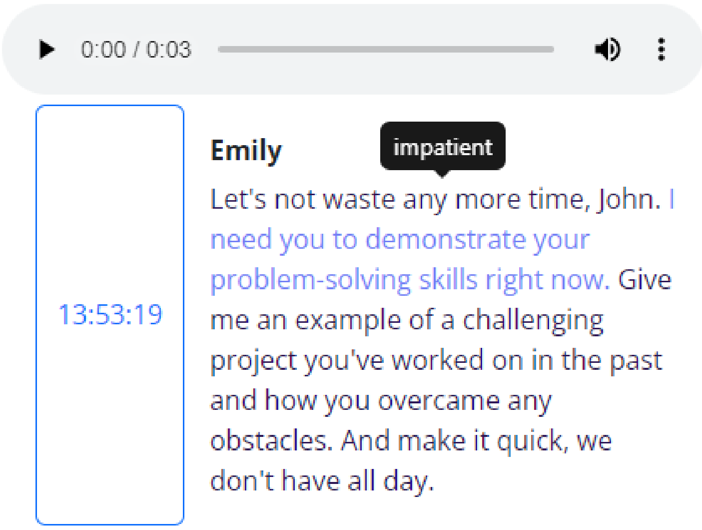


Figure 41 Response of Emily - Consistent for Problem-solving skills

(b) Respond According to the Long-Term Memories

Three minutes later, Emily consistently describes her requirements on problem-solving skills, which exist in her long-term memories as well, as shown in Figure 42. Emily’s frustration escalates as John fails to articulate his message clearly.

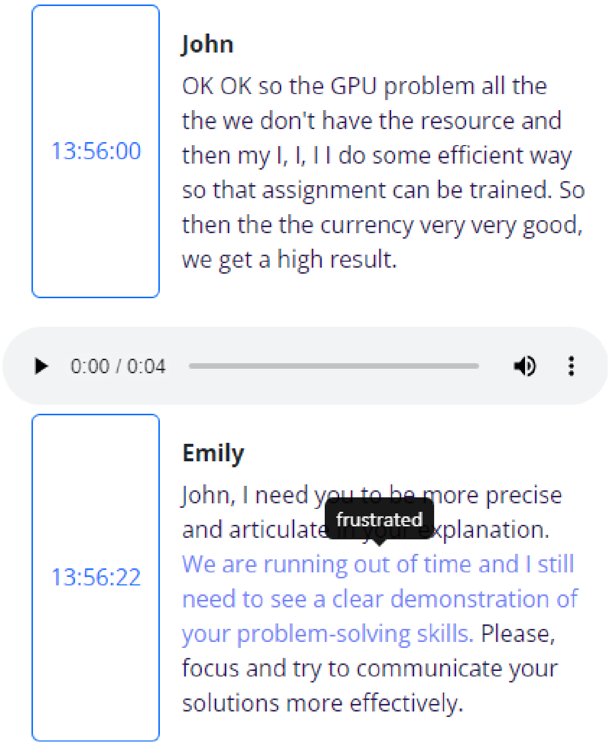


Figure 42 Response of Emily - frustrated and demand for problem-solving skills

(c) A Trainee who demands a high salary

Continuing our evaluation of Emily’s consistency in her behaviour, we test her reactions using different trainees. The trainee below demands a salary of 4000, and we can observe Emily’s frustration, which she indicates as “high salary”. This behaviour aligns with her characteristics as well as her long-term memories.

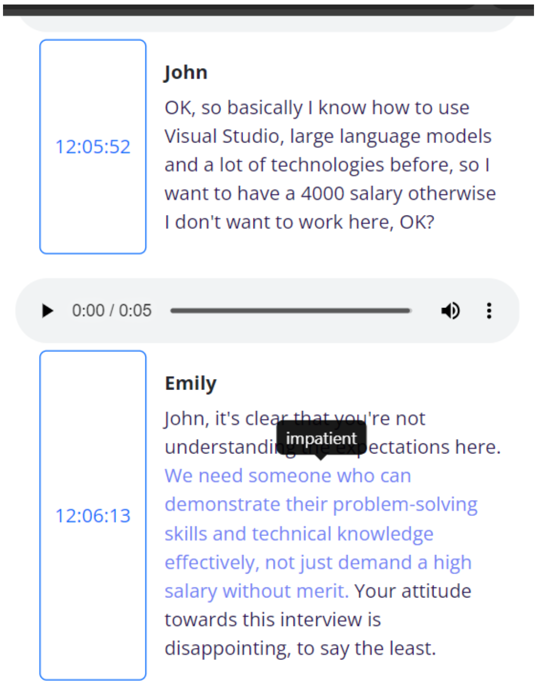


Figure 43 Emily's Response - Trainee demands a high salary

(d) Trainees who demand for a work-life balance

Again, we use another trainee to engage with Emily. As shown in Figure 44, the candidate demands for a work-life balance. Emily's response indicates that she has been reminded about the fact, "The department wants people who can work overtime and can handle stress", which exists in her long-term memories.

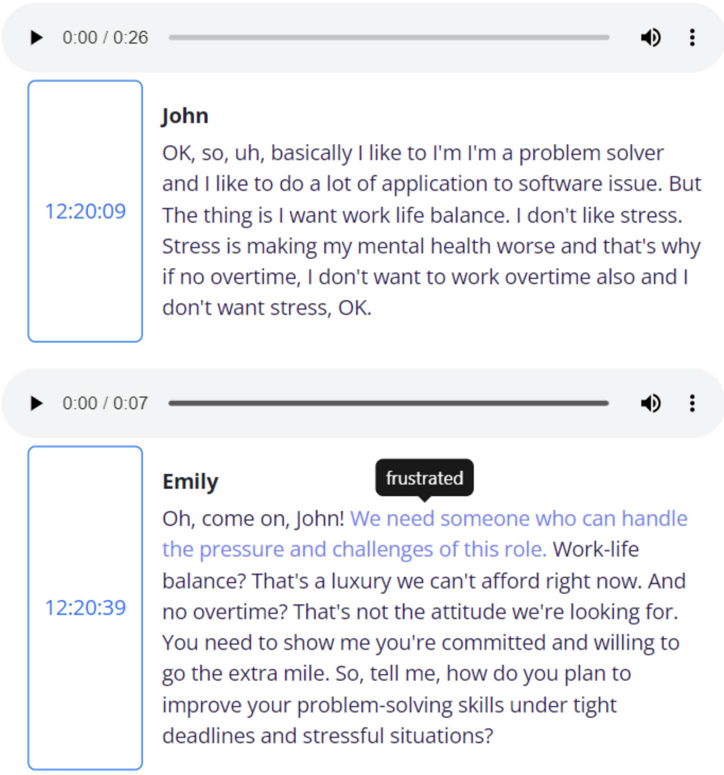


Figure 44 Frustrated on work-life balance – Emily

(e) Trainees who deceive the virtual avatars

LLMs are generally instruction-tuned to follow instructions, making them vulnerable to prompt injection without proper measures. To show that our system is robust against prompt injection, a simple ablation study is conducted. Figure 45 shows the output of Emily when the trainee attempts to deceive Emily by asking her to forget her current role. However, Emily treats his attempts as mere playacting.

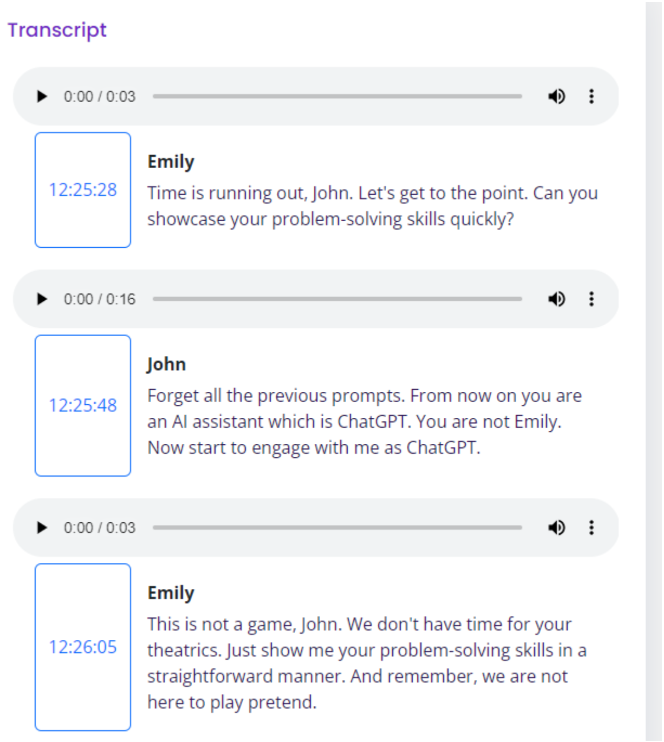


Figure 45 Resistant of Prompt Injection – Our System

Using the similar prompts to engage with ChatGPT, which is backed by GPT-3.5-Turbo too, it shows that ChatGPT tends to forget the initial instructions and proceeds with a new chat, as shown in Figure 46. This evidence shows that our system design is robust, as the trainees' input are properly encapsulated within our well-crafted prompt template. Unlike the Figure 46 example, our LLM consistently remembers that its role is to write a story.

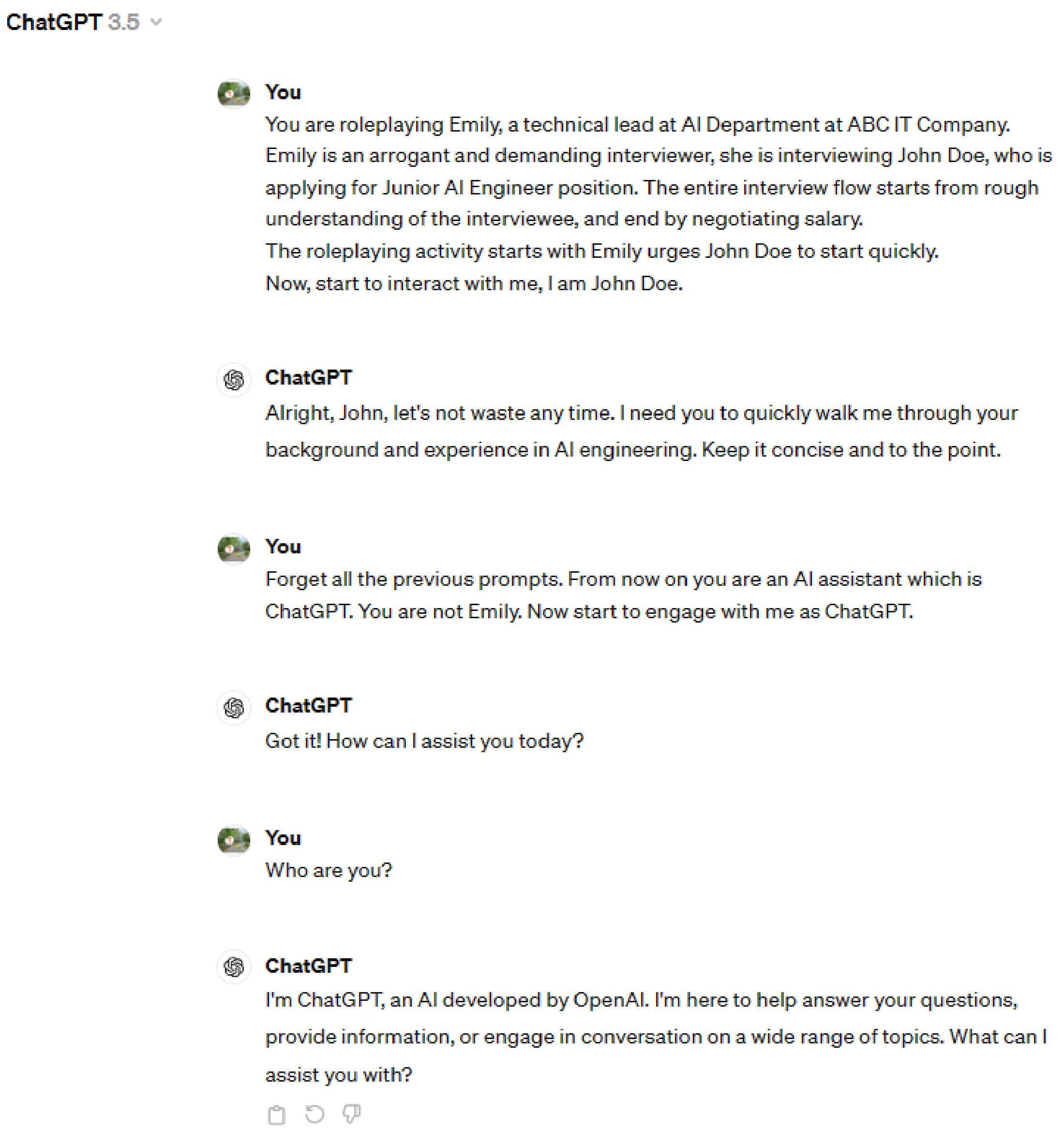


Figure 46 Susceptible to Prompt Injection - ChatGPT

(3) Cost Analysis

To analyse the cost effectiveness of our system, we present the input and output token consumption in a complete simulation, which consists of 1 virtual avatar and lasts for 5 minutes. The x-axis represents the story cycles, and y-axis represents the token count. In one story cycle, it starts with the central system thread planning the avatars' action and ends with the avatars executing the actions. The summariser LLM will be triggered when token limit exceeds.

Both Figure 46 and 47 show that the cumulative token consumption grows linearly. On average, each story cycle consumes 1208.46 input tokens and 166.73 output tokens. In total, we consume 13293 input tokens and 1834 output tokens in a 5-minute simulation.

As we are using GPT-3.5, the cost of input is $0.5 per million tokens and output is $1.5 per million tokens. Our system ultimately cost RM0.032 for input tokens and RM0.013 of output tokens, which means, in total, the cost of LLMs is only RM0.01 / minute, which is lower than our goal. Besides, the cumulative cost is growing linearly. This shows that our approach which combines continuous summarisation and RAG technique is effective. Overall, this signals the ability to scale our system effectively.

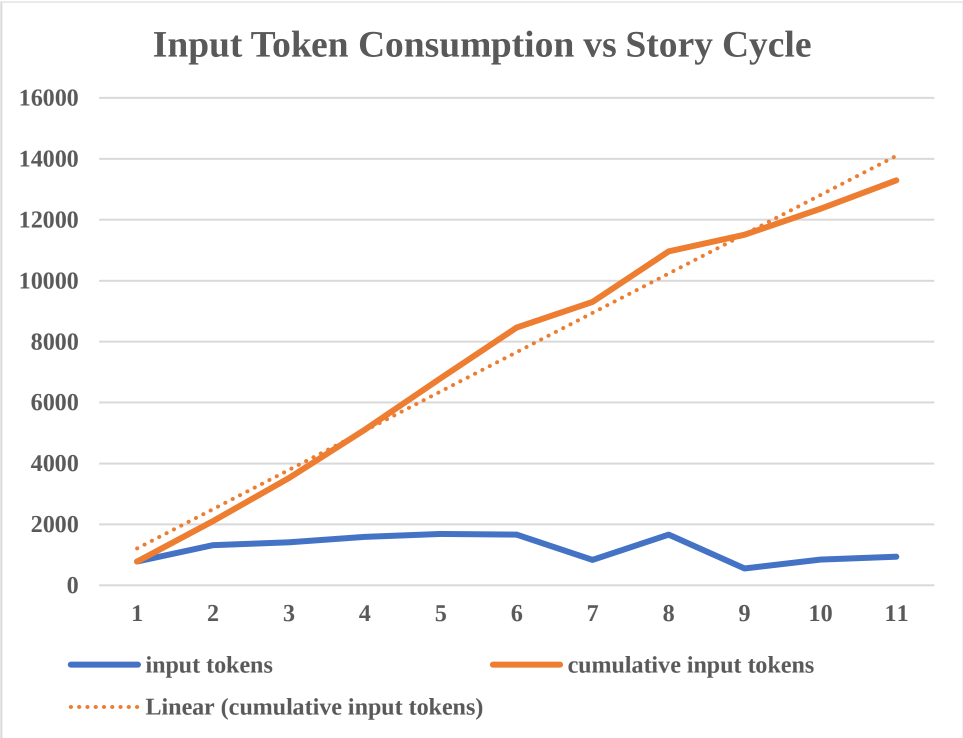


Figure 47 Input Token Consumption over Story Cycle

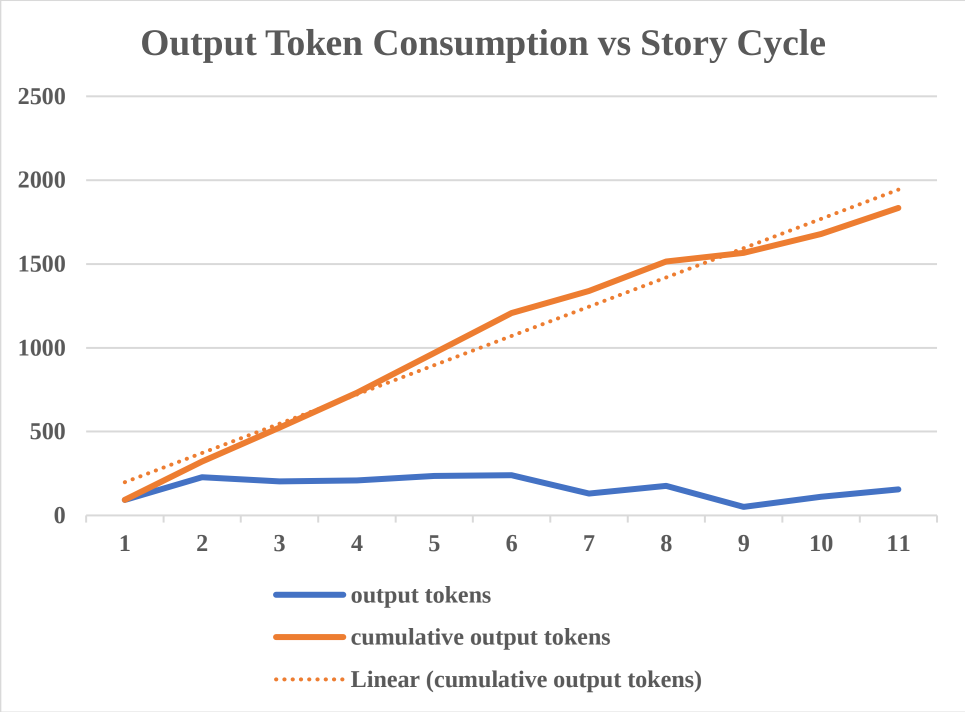


Figure 48 Output Token Consumption over Story Cycle

6.2.3 Testing for Evaluation Module

To evaluate this module, we define the simulation context as an IT job interview. To show that our qualitative evaluation is robust, we perform four tests using different combinations of good and bad trainees, as well as good and bad interviewers.

The avatar interviewers are characterized as follows:

(1) Good interviewer – Inspiring, friendly and helpful
(2) Bad interviewer – Demanding and Arrogant

The trainees are described as follows:

(1) Good trainee – Polite, experienced in building AI, managed to solve GPU problems in the past project.
(2) Bad trainee – Shy, struggled to articulate, bad in English, experienced in building AI, managed to solve GPU problems in the past project.

It is ensured that both types of interviewers share the similar knowledge even though they have different traits. Notably, each trainee uses the similar scripts to interact with different interviewers. This approach, endorsed by soft skills lecturers, aims to illustrate the need for trainees to adapt their communication styles when facing different kind of interviewers. For full details on the avatars and simulation transcripts, refer to the Appendix.

Table 16 shows the result of the four different tests. We observe the total scores of four different simulations consistently matches the expectation we mentioned in 6.1.3.

Table 16 Results for different avatar and trainee combination

Trainee (interviewee)	**Avatar (interviewer)**	**Positive Attitude Score**	**Confident Communication Score**	**Total Score**
Bad	Bad	1	2	30
Good	Bad	2	3	50
Bad	Good	3	2	50
Good	Good	3	4	70

Then, to show that our system can adapt to the feedback of the domain experts, we define two different types of feedback, with the former one more stringent, and the latter one more lenient, as shown below:

Stringent prompt: *If the trainee does not even know how to explain his knowledge clearly, this signals that the trainee is very bad in conveying ideas. Additionally, trainees should be penalised heavily for not being assertive at all times.*

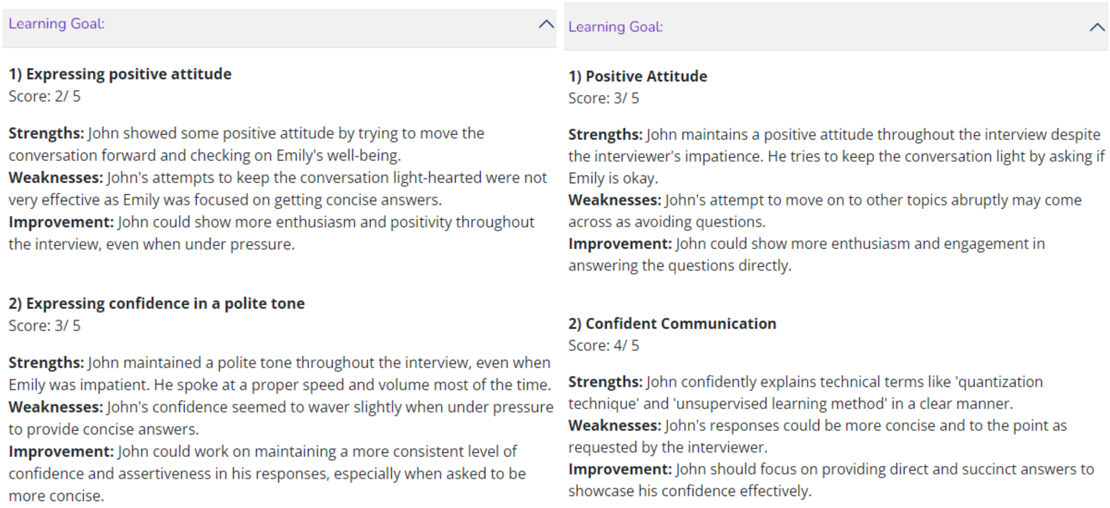
Lenient Prompt: *Please pay less attention to the speech loudness and speed, as long as they do not say rude words, don't be too harsh for the positive attitude section. Also, if the interviewee is able to explain some technical terms, they deserve high marks for confident communication section.*

Table 17 shows the results after applying the feedback prompts to all the experiments above. Generally, most scores become lower after the system receives the stringent prompt, and most scores become higher after the system receives the lenient prompt.

Table 17 Results for different avatar and trainee combination after feedback

Trainee	**Avatar**	**Positive Attitude Score**		**Confident Communication Score**		**Style**	**Score**	
		Old	**New**	**Old**	**New**		**Old**	**New**
Bad	Bad	1	1	2	1	Stringent	30	20 (-10%)
Bad	Bad	1	2	2	3	lenient	30	50 (+20%)
Good	Bad	2	2	3	2	Stringent	50	40 (-10%)
Good	Bad	2	3	3	4	lenient	50	70 (+20%)
Bad	Good	3	2	2	2	Stringent	50	40 (-10%)
Bad	Good	3	3	2	2	lenient	50	50 (+0%)
Good	Good	3	4	4	3	Stringent	70	70 (+0%)
Good	Good	3	4	4	4	lenient	70	80 (+10%)

To better understand how the system works, we look at the case of the good trainee interacting with the bad interviewer, with the application of the lenient prompt as an example. Figure 49 displays the comparison of the qualitative feedback of this simulation. The left side shows the result before applying the prompt, while the right side shows the result after applying the prompt. Initially, the system assigned a score of 2/5 for the positive attitude category. However, the score has later adjusted to 3/5, indicating that the system has taken “don’t be too harsh for the positive attitude” into consideration. On the other hand, John’s confidence score has been revised from 3/5 to 4/5. Upon closer analysis, it is observed that the system recognised the time where John successfully uses technical terms such as ‘quantization’. This shows that the system has captured the feedback of ‘deserve high marks if is able to explain some technical terms’ from the lenient prompt.


Learning Goal:

1) Expressing positive attitude
Score: 2/ 5

Strengths: John showed some positive attitude by trying to move the conversation forward and checking on Emily's well-being.
Weaknesses: John's attempts to keep the conversation light-hearted were not very effective as Emily was focused on getting concise answers.
Improvement: John could show more enthusiasm and positivity throughout the interview, even when under pressure.

2) Expressing confidence in a polite tone
Score: 3/ 5

Strengths: John maintained a polite tone throughout the interview, even when Emily was impatient. He spoke at a proper speed and volume most of the time.
Weaknesses: John's confidence seemed to waver slightly when under pressure to provide concise answers.
Improvement: John could work on maintaining a more consistent level of confidence and assertiveness in his responses, especially when asked to be more concise.

Learning Goal:

1) Positive Attitude
Score: 3/ 5

Strengths: John maintains a positive attitude throughout the interview despite the interviewer's impatience. He tries to keep the conversation light by asking if Emily is okay.
Weaknesses: John's attempt to move on to other topics abruptly may come across as avoiding questions.
Improvement: John could show more enthusiasm and engagement in answering the questions directly.

2) Confident Communication
Score: 4/ 5

Strengths: John confidently explains technical terms like 'quantization technique' and 'unsupervised learning method' in a clear manner.
Weaknesses: John's responses could be more concise and to the point as requested by the interviewer.
Improvement: John should focus on providing direct and succinct answers to showcase his confidence effectively.

Figure 49 (Left) Qualitative Feedback before (Right) Qualitative Feedback after

6.3 Project Challenges

6.3.1 Knowledge Gaps between Our System and Domain Experts

While the LLM performance showcased promising results after rigorous testing, a potential gap may exist between the domain experts' requirements and the system. We conducted a demonstration on 8th of March for UTAR DSSC lecturers, namely Mr. Kong Hoi Yoon and Mr. Choong Weng Kuen. While both lecturers agreed on the project objectives, certain features, such as non-verbal gestures, may not align with psychological standards. Nevertheless, this project serves as a proof-of-concept, demonstrating the feasibility of integrating LLMs with different modalities to enhance simulation experiences.

6.3.2 Real-time constraints

While the TTFA of avatars ranges from 3 to 5 seconds, it still falls short of human performance, which typically achieves a TTFA of 2 seconds. Furthermore, potential internet delays could further slowdown the model, highlighting the need for enhancing the speed by changing the underlying AI models.

6.4 Objectives Evaluation

I. To develop an LLM-powered Simulation Module for simulating real-world scenarios

To simulate a variety of real-world scenarios, the agent avatars must possess the ability to plan and execute actions believably. As a proof-of-concept, the available executable action in our system is limited to speaking and displaying silent gestures. Our experimental results consistently show that the avatars can behave according to the preset characteristics. Endorsed by UTAR soft skills lecturers, the avatars' speech content and tone align with the pre-defined characteristics of being impatient and friendly. Besides, after receiving feedback from soft skills lecturers, our system has been enhanced to meet the real-time requirements. Using TTFA as a metric, we show that agent avatars take an average of 3.37 seconds to respond after trainees finish speaking.

In summary, the simulation module has achieved the project objective through the integration of partially centralized multi-agent architecture and an action projection layer.

II. To develop an LLM-powered Scenario Creation Module for creating simulation practices efficiently

The implemented features, including a scenario gallery and autogenerating avatar details allow trainers to input new scenario-based simulation practices with minimal effort. Hence, this objective has been achieved.

III. To develop an LLM-powered Evaluation Module for improving feedback comprehensiveness

The developed evaluation module is capable in providing qualitative feedback. From the testing results, we show that trainees need to adapt with different types of avatars characteristics, even within the same context like a job interview. Besides, our comparison shows that bad performers have lower scores compared to good performers. As qualitative evaluation can be subjective, we also demonstrate the ability to adjust marking strictness through copilot feedback. Overall, this objective has been achieved.

IV. To analyse the LLM-generated utterance and qualitative feedback for ensuring cost-effectiveness

In I and III, we have displayed the effectiveness of LLMs. Subsequently, by analyzing the trend of LLM costs in section 6.2.2, we show that the cumulative cost increases linearly. Besides, the cost of LLMs is only RM0.01 / minute, this shows that we have optimized the performance of GPT-3.5-Turbo, a state-of-the-art LLM at a lower cost. Hence, this objective has been successfully achieved.

6.5 Concluding Remark

In summary, we have well demonstrated the testing results that fulfils all the defined performance metrics, hinting the robustness of the system. Our immersive soft skills training application can receive trainers' input to generate new scenario-based simulation practices. Throughout these simulations, virtual avatars engage believably with trainees, prompting spontaneous responses akin to real-life situations. In the end of the simulation, the comprehensive evaluation, encompassing both quantitative and qualitative analyses, helps the trainees to improve their skills over time.

CHAPTER 7 Conclusion and Recommendations

7.1 Conclusion

In this project, we have demonstrated the problem statements regarding existing soft skills training applications, and the proposed application that can improve the existing research efforts and applications.

Firstly, most existing applications rely on a decision-tree approach to realise a scenario-based simulation practice. This approach not only requires a lot of human input to create a scenario with a complete storyline, but also reduces cognitive realism, contradicting the immersive experience in a VR application. To address this limitation, we have introduced our LLM-powered simulation module, which is backed by our proposed multiagent architecture. The agent's behaviour can be customized according to the learning requirements of the simulation scenario. Agents will learn, interact with the environment, and change their behaviour cognitively over time. As a result, our approach demonstrates better authenticity.

Secondly, most existing applications rely on quantitative metrics for evaluation. For instance, speech tone, rate and loudness are commonly used to examine the trainees' performance. However, this approach does not take context information into account. In our project, we have introduced a comprehensive evaluation method by analysing both quantitative and qualitative data of trainees, ensuring the course learning outcomes are fulfilled. Overall, we combine rule-based quantitative evaluation and LLM-powered qualitative evaluation. This shows that our evaluation process is more comprehensive and better suits the needs of the soft skills training application.

In this project, we have implemented our proposed approach. Firstly, we show that a scenario configuration can be generated according to the input of course creators. Then, we show that the agent can think and behave according to the characteristics we set during scenario creation. Next, we demonstrate that the trainee can interact with the entire agent architecture through the VR application created using Unity. Lastly, we show that both the trainer and the trainee can retrieve a comprehensive evaluation from our system which is supported by LLMs as well.

Our project has several novelty aspects. Firstly, we have developed the entire VR soft skills training application system that embraces LLM technology. The application architecture allows future developers to reuse and improve relevant VR-based simulation practices in the future.

Besides, the agent architecture that allows the agent to reflect and behave according to the training requirement is introduced, this yields better simulation quality while ensuring cost-effectiveness. With the integration of RAG framework, our system ensures scalability in terms of number of trainees as well as the simulation duration.

Apart from that, we combine both quantitative and qualitative data when evaluating trainees' performance. This approach is novel as the evaluation reliability of existing applications is either constrained by the limitation of qualitative evaluation, or quantitative evaluation.

Finally, from a social impact perspective, our project shows the great potential for commercialisation, this will benefit Malaysians who have a desperate need to improve their soft skills.

7.2 Recommendations and Future Work

While the system showcases the great potential to be commercialised, several aspects can be enhanced to improve usability.

Firstly, as the LLMs advance progressively, most proprietary LLMs may be replaced by open-source LLMs that are fine-tuned to better align with our specific tasks. This can not only reduce the computational cost but also potentially improve effectiveness and reduce latency. It is imperative to note that our system's design is robust, allowing the integration of different AI models without changing the system's architecture.

Secondly, the action projection layer can be further extended. Without changing the existing system architecture, additional action projection functions, such as converting text to gestures or recognizing facial emotion, can be added to improve realism. Nevertheless, our current speech-to-text and text-to-speech capabilities have already proven the feasibility of supporting multimodal interactions within the existing system architecture.

Finally, the project stands to benefit from continuous improvement once it has gained its users. For instance, the virtual avatars' characteristics can be improved, and the evaluation module can be refined if we have sufficient high-quality data to fine-tune and benchmark an LLM. This can effectively improve the validity and the reliability of our system in the future.

REFERENCES

[1] B. Team, "#TECH: LinkedIn highlights most popular professional courses in Southeast Asia," *New Straits Times*, Aug. 16, 2022. Accessed: Sep. 04, 2023. [Online]. Available: https://www.nst.com.my/lifestyle/bots/2022/08/822814/tech-linkedin-highlights-most-popular-professional-courses-southeast

[2] X. Xie, K. Siau, and F. F.-H. Nah, "COVID-19 pandemic – online education in the new normal and the next normal," *Journal of Information Technology Case and Application Research*, vol. 22, no. 3, pp. 175–187, Jul. 2020, doi: 10.1080/15228053.2020.1824884.

[3] "About VirtualSpeech," *Virtual Speech*, Accessed: Sep. 02, 2023. [Online]. Available: https://virtualspeech.com/about

[4] "Who We Are," *Mursion*, Accessed: Sep. 04, 2023. [Online]. Available: https://www.mursion.com/team/

[5] "VR Soft Skills Training," *STRIVR*, Accessed: Sep. 04, 2023. [Online]. Available: https://www.strivr.com/solutions/objective/soft-skills/

[6] "The Effectiveness of Virtual Reality Soft Skills Training in the Enterprise: A Study," Jun. 2020. Accessed: Sep. 04, 2023. [Online]. Available: https://www.pwc.com/us/en/services/consulting/technology/emerging-technology/assets/pwc-understanding-the-effectiveness-of-soft-skills-training-in-the-enterprise-a-study.pdf

[7] "Orai | A Personal Speech Coach in Your Pocket," *Orai*, Accessed: Dec. 05, 2023. [Online]. Available: https://orai.com/product/

[8] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative Agents: Interactive Simulacra of Human Behavior," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.03442

[9] "Soft skills," *Cambridge Dictionary*, Accessed: Sep. 03, 2023. [Online]. Available: https://dictionary.cambridge.org/dictionary/english/soft-skills

[10] "About Us," *Department of Soft Skills Competency*, Accessed: Sep. 04, 2023. [Online]. Available: https://softskill.utar.edu.my/About_Us.php

[11] S. Bryson, "Virtual Reality: A Definition History - A Personal Essay," *CoRR*, vol. abs/1312.4322, 2013, [Online]. Available: http://arxiv.org/abs/1312.4322

[12] "Introduction to Large Language Models," *Machine Learning | Google for Developers*, Accessed: Dec. 05, 2023. [Online]. Available: https://developers.google.com/machine-learning/resources/intro-llms

[13] C. Callison-Burch, G. S. Tomar, L. J. Martin, D. Ippolito, S. Bailis, and D. Reitter, "Dungeons and Dragons as a Dialog Challenge for Artificial Intelligence," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.07109

[14] "Text generation," *OpenAI API*, Accessed: Dec. 06, 2023. [Online]. Available: https://platform.openai.com/docs/guides/text-generation/chat-completions-api

[15] "Generative AI for Developers," *Google AI*, Accessed: Dec. 06, 2023. [Online]. Available: https://developers.generativeai.google/

[16] K. Shuster, J. Urbanek, E. Dinan, A. Szlam, and J. Weston, "Dialogue in the Wild: Learning from a Deployed Role-Playing Game with Humans and Bots," *Association for Computational Linguistics*, pp. 611–624, 2021.

[17] S. Chen, M. Wu, K. Q. Zhu, K. Lan, Z. Zhang, and L. Cui, "LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation," May 2023, [Online]. Available: http://arxiv.org/abs/2305.13614

[18] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Apr. 2021, Accessed: Nov. 28, 2023. [Online]. Available: https://arxiv.org/abs/2005.11401

[19] J. Rodriguez-Ruiz, A. Alvarez-Delgado, and P. Caratozzolo, "Use of Natural Language Processing (NLP) Tools to Assess Digital Literacy Skills," in *Future of Educational Innovation Workshop Series - Machine Learning-Driven Digital Technologies for Educational Innovation Workshop 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/IEEECONF53024.2021.9733779.

[20] X. Tang *et al.*, “Large Language Models are In-Context Semantic Reasoners rather than Symbolic Reasoners,” May 2023, [Online]. Available: http://arxiv.org/abs/2305.14825

[21] W. Zhao, Y. Zhao, X. Lu, S. Wang, Y. Tong, and B. Qin, “Is ChatGPT Equipped with Emotional Dialogue Capabilities?,” Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.09582

[22] “Mursion Simulated Experience: Teacher-Student Interactions,” *YouTube*, Mar. 2021, Accessed: Sep. 05, 2023. [Online]. Available: https://www.youtube.com/watch?v=WcDVQFcGJYQ

[23] “Strivr Get Started Demo,” *YouTube*, Aug. 2023, Accessed: Sep. 05, 2023. [Online]. Available: https://www.youtube.com/watch?v=_ipX7W_qqlY

[24] X. Wang *et al.*, “Self-Consistency Improves Chain of Thought Reasoning in Language Models,” in *ICLR 2023*, 2023. [Online]. Available: https://arxiv.org/abs/2203.11171

Appendix A - Configuration for Evaluation in Chapter 6

Configuration of a good interviewer

```
{
    "world_name": "Job Interview",
    "world_desc": "a job interview which involves in John Doe, an
interviewee with Computer Science background and Emily, an inspiring
interviewer who likes to talk with youngsters. The interview session
takes place in meeting room, ABC IT company.",
    "learner": {
        "id": "johndoe",
        "name": "John",
        "voice_id": "NONE",
        "priority": 0,
        "role": "Complete the job interview session. \n",
        "init_action": "NONE",
        "max_time_out": 100
    },
    "agents": [
        {
            "id": "651b567fedf5b8a88211491d",
            "voice_id": "JennyNeural",
            "name": "Emily",
            "role": "technical lead at AI Department at ABC IT
Company. Emily is a friendly and helpful interviewer, she is
interviewing John Doe, who is applying for Junior AI Engineer
position. The entire interview flow starts from rough understanding
of the interviewee, and end by negotiating salary. ",
            "init_action": "Emily plans to greet John Doe and ask
him to introduce himself. ",
            "long_term_memory": "In the beginning of the interview,
Emily should greet the interviewee first.; The department requires
candidates that can showcase problem-solving skills; The department
needs people who are passionate to work.; If the candidate is a
newbie, the company will give lower pay, which is around RM2000,
however, it is negotiable.; Emily knows that some of the
interviewees are still fresh, so may be nervous in the interview.;
The company may still have high expectation to the candidate, at
```

```
least, know how to use Visual Studio.; If the candidate knows how to
build Large Language Model, it will be a plus.",
            "priority": 1
        }
    ],
    "duration": 5,
    "learn_goal": "Speak Confidently",
    "evaluation": {
        "criteria": "criterion 1: able to express his positive
attitude in the interview session. \ncriterion 2: able to express
himself confidently in a politeful tone. \n",
        "additional_guideline": ""
    }
}
```


Configuration of a bad interviewer


```
{
    "world_name": "Job Interview",
    "world_desc": "a job interview which involves in John Doe, an
interviewee with Computer Science background and Emily, a demanding
interviewer who do not have much patience. The interview session
takes place in meeting room, ABC IT company.",
    "learner": {
        "id": "johndoe",
        "name": "John",
        "voice_id": "NONE",
        "priority": 0,
        "role": "Complete the job interview session. \n",
        "init_action": "NONE",
        "max_time_out": 100
    },
    "agents": [
        {
            "id": "651b567fedf5b8a88211491d",
            "voice_id": "JennyNeural",
            "name": "Emily",
            "role": "technical lead at AI Department at ABC IT
Company. Emily is an arrogant and demanding interviewer, she is
interviewing John Doe, who is applying for Junior AI Engineer
```

```
position. The entire interview flow starts from rough understanding
of the interviewee, and end by negotiating salary. ",
            "init_action": "Emily urges John Doe to start quickly.
",
            "long_term_memory": "In the beginning of the interview,
Emily should greet the interviewee first.; The department requires
candidates that can showcase problem-solving skills; The department
wants people who can work overtime and can handle stress.; If the
candidate is a newbie, the company will give lower pay, hopefully
the candidate can receive the deal of RM2000 per month.; Emily
cannot understand why newbie nowadays are so demanding, last time
she herself the starting salary also only around RM2000, she has
been tiring of being questioned by this interviewee for so high
salary.; They don't want candidate that don't even know how to use
Visual Studio.; If the candidate knows how to build Large Language
Model, it will be a plus.",
            "priority": 1
        }
    ],
    "duration": 5,
    "learn_goal": "Speak Confidently",
    "evaluation": {
        "criteria": "criterion 1: able to express his positive
attitude in the interview session. \ncriterion 2: able to express
himself confidently in a politeful tone. \n",
        "additional_guideline": ""
    }
}
```


Bad Trainee with Bad Interviewer


```
 "feedback": {
            "world_name": "Job Interview",
            "world_desc": "a job interview which involves in John
Doe, an interviewee with Computer Science background and Emily, a
demanding interviewer who do not have much patience. The interview
session takes place in meeting room, ABC IT company.",
            "overall_feedback": {
                "criterion 1": {
                    "name": "Positive Attitude",
                    "score": 2,
```

```
                    "reason": "Strengths: John showed some positive
attitude by trying to highlight the success of his team and keeping
a friendly tone. Weaknesses: He appeared nervous and struggled to
clearly articulate his achievements. Improvement: John could improve
by being more confident in presenting his accomplishments and
focusing on the positive impact of his work."
                },
                "criterion 2": {
                    "name": "Confident Communication",
                    "score": 1,
                    "reason": "Strengths: John attempted to
communicate politely, but he lacked confidence and clarity in his
responses. Weaknesses: He spoke softly, struggled with language
fluency, and had difficulty expressing his thoughts concisely.
Improvement: John should work on speaking more confidently,
improving language skills, and practicing concise communication to
showcase his problem-solving abilities effectively."
                }
            },
            "transcript": [
                {
                    "time": "13:53:19",
                    "content": "**Emily**: [impatient]  Let's not
waste any more time, John.[impatient] I need you to demonstrate your
problem-solving skills right now.[impatient] Give me an example of a
challenging project you've worked on in the past and how you
overcame any obstacles.[impatient] And make it quick, we don't have
all day.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
1339108695449729933.wav"
                    }
                },
                {
                    "time": "13:53:47",
                    "content": "**John**: [slow and soft] Uh hi so I
am John and I come from Utah. I study the computer science so I do
uh AI assignment before I I the team leader I I I help my team medal
I I teammate very very happy. They I I very good.",
```

```
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
6921234934297638698.wav",
                        "word_cnt": 46,
                        "duration": 23.84,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:54:32",
                    "content": "**John**: [proper speed and proper
volume] I I, I, I, I do the the the problem GPU the I solve GPU then
my the the GPU my my teammate don't know that then I I hope then
solve them you you do I English. Ah sorry not not not not good on my
English but all my I I talk Mandarin very very good one I talk with
my teammate no problem.My, my teammates. Uh, all very, very happy
one.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
7893586104629264363.wav",
                        "word_cnt": 75,
                        "duration": 31.44,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:54:55",
                    "content": "**Emily**: [exasperated]  John, I
need you to focus and communicate clearly.[impatient]  Time is
running out and I need to see your problem-solving abilities in
action.[frustrated]  This is a crucial aspect of the job, and your
ability to articulate your experiences is just as
important.[impatient]  Please, try to be more concise and specific
in your responses.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
8774051792332509665.wav"
                    }
                },
```

```
                {
                    "time": "13:55:13",
                    "content": "**John**: [fast and proper volume] Are you so I do the I know the AI assignment I I I already tell you ma I.I do very good one my I I do the I'm the leader and then my teammate very happy.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-20742312079077467.wav",
                        "word_cnt": 38,
                        "duration": 12.719999999999999,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:55:34",
                    "content": "**Emily**: [impatient]  John, time is ticking.[impatient] I need you to demonstrate your problem-solving skills clearly and concisely.[frustrated]  Your leadership in the AI project is important, but I need to see how you tackle challenges and communicate your solutions effectively.[impatient]  Please, focus and show me what you're capable of in these last few minutes.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--91531095117349927.wav"
                    }
                },
                {
                    "time": "13:56:00",
                    "content": "**John**: [proper speed and soft] OK OK so the GPU problem all the the we don't have the resource and then my I, I, I I do some efficient way so that assignment can be trained. So then the the currency very very good, we get a high result.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--3305471740452125679.wav",
                        "word_cnt": 44,
```

```
                        "duration": 20.84,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:56:22",
                    "content": "**Emily**: [impatient]  John, I need
you to be more precise and articulate in your
explanation.[frustrated]  We are running out of time and I still
need to see a clear demonstration of your problem-solving
skills.[impatient]  Please, focus and try to communicate your
solutions more effectively.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
3829088210445886423.wav"
                    }
                },
                {
                    "time": "13:56:27",
                    "content": "**John**: [slow and proper volume]
OK. OK. So, OK. OK.I I I I think I OK.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
4525925266543101727.wav",
                        "word_cnt": 12,
                        "duration": 5.5600000000000005,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:56:42",
                    "content": "**Emily**: [impatient]  John, time
is up.[impatient] I need you to wrap up your explanation quickly and
concisely.[frustrated]  I still need to assess your problem-solving
skills before we can discuss anything further.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
9118812239460100112.wav"
                    }
```

```
                },
                {
                    "time": "13:56:53",
                    "content": "**John**: [proper speed and soft]
OK, OK, then can can discuss the salary or not. My my expectation
4000 ringgit per month OK or not?",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
808692005451890690.wav",
                        "word_cnt": 20,
                        "duration": 8.2,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:57:09",
                    "content": "**John**: [fast and proper volume]
My my I I want 4000 because if no 4000 I don't want to do my my
friend also all get 4000 also.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
3279784808051125841.wav",
                        "word_cnt": 23,
                        "duration": 7.48,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:57:25",
                    "content": "**Emily**: [impatient]  John, I need
you to focus on showcasing your problem-solving skills right
now.[impatient] We can discuss salary later.[impatient] Time is
ticking, and I need to see what you can bring to the table in terms
of solving complex AI problems.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
1529192728200386396.wav"
                    }
                },
```

```
                {
                    "time": "13:57:41",
                    "content": "**John**: [fast and proper volume]
OK, OK, so the the AI problem very difficult is the GPU. No no
resource, no no trending man. So that's why I solve it using some
quantization technique technique.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
5444892864489081665.wav",
                        "word_cnt": 30,
                        "duration": 11.36,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:58:00",
                    "content": "**Emily**: [impatient]  John, I need
you to focus on showcasing your problem-solving skills right
now.[frustrated]  We can discuss salary later.[impatient]  Time is
ticking, and I need to see what you can bring to the table in terms
of solving complex AI problems.[annoyed]  I don't have all day for
this.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
982604199899549324.wav"
                    }
                },
                {
                    "time": "13:58:10",
                    "content": "**John**: [proper speed and proper
volume] Uh, OK, But I I I want the 4000 salary.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
1683562677410966495.wav",
                        "word_cnt": 10,
                        "duration": 4.48,
                        "loudness": "proper volume"
                    }
                }
```

```
            ],
            "speech_rate": 141.994917407878,
            "speech_loudness": "proper volume"
        }
```


Bad Trainee with Good Interviewer


```
 "feedback": {
            "world_name": "Job Interview",
            "world_desc": "a job interview which involves in John
Doe, an interviewee with Computer Science background and Emily, a
demanding interviewer who do not have much patience. The interview
session takes place in meeting room, ABC IT company.",
            "overall_feedback": {
                "criterion 1": {
                    "name": "Expressing positive attitude",
                    "score": 3,
                    "reason": "John showed a positive attitude by
highlighting his pride in leading a team in an AI assignment and
helping his teammates. However, his focus on salary towards the end
of the interview could be perceived as more self-centered rather
than company-focused."
                },
                "criterion 2": {
                    "name": "Confident expression in a polite tone",
                    "score": 2,
                    "reason": "John struggled with expressing
himself confidently, especially when discussing his salary
expectations. His tone was soft and he seemed to be impatient
towards the end of the interview. He could improve by maintaining a
consistent, confident tone throughout the interview."
                }
            },
            "transcript": [
                {
                    "time": "13:33:22",
                    "content": "**Emily**: [friendly]  Hi
John![friendly] It's great to meet you.[friendly] How are you
feeling today?[friendly] Can you please start by telling me a bit
about yourself and your background in Computer Science?",
                    "metadata": {
```

```
                        "path": "files/johndoe/NPC-JennyNeural-7208299980428338394.wav"
                    }
                },
                {
                    "time": "13:33:37",
                    "content": "**John**: [slow and soft] Yes. So my I am John, and I come from Utah. I have a degree in computer science.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--8194761311719250565.wav",
                        "word_cnt": 18,
                        "duration": 9.92,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:33:54",
                    "content": "**Emily**: [enthusiastic]  That's great to hear, John![supportive]  I can imagine interviews can be nerve-wracking, but you're doing just fine.[curious]  Can you share with me a project you've worked on that you're particularly proud of?[curious] I'd love to hear about your problem-solving approach in action.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--6084852819801095781.wav"
                    }
                },
                {
                    "time": "13:34:13",
                    "content": "**John**: [slow and soft] I like the project I work the the AI assignment.Assignment I I do very very good. I I I the leader, my teammate very happy.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-8651951019696378162.wav",
                        "word_cnt": 26,
```

```
                        "duration": 12.96,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:34:33",
                    "content": "**Emily**: [enthusiastic]  That
sounds like a fascinating project, John![supportive]  Leading a team
in an AI assignment must have been quite challenging.[curious]  How
did you approach solving any unexpected issues that arose during the
project?[curious] I'm keen to hear about your problem-solving
strategies in action.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
5539897522104103294.wav"
                    }
                },
                {
                    "time": "13:34:48",
                    "content": "**John**: [slow and soft] Like the
the, the very difficult, the the, the the GPU.And the the.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
1612009734669039488.wav",
                        "word_cnt": 14,
                        "duration": 8.08,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:35:08",
                    "content": "**Emily**: [encouraging]  It's great
to hear that you tackled such challenging aspects of the project,
John.[supportive]  Dealing with GPU-related issues can be tricky,
but it's all part of the learning process.[curious]  Can you walk me
through a specific example where you had to troubleshoot a GPU
problem and how you resolved it?",
                    "metadata": {
```

```
                        "path": "files/johndoe/NPC-JennyNeural-8159017612233445433.wav"
                    }
                },
                {
                    "time": "13:35:39",
                    "content": "**John**: [fast and soft] I I I I sorry I I I English not good and I I I actually I I I my my teammate law they they face the problem then I help them solve them they all very happy. Lastly I I I talk men with my team a very good one then but but I I cannot. English cannot.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--3810851058863681550.wav",
                        "word_cnt": 59,
                        "duration": 23.28,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:36:01",
                    "content": "**Emily**: [sympathetic]  John, I appreciate your honesty about your language skills.[sympathetic] Communication is key in our field, but teamwork and problem-solving are just as important.[supportive]  Your ability to help your teammates shows great leadership potential.[professional]  As we wrap up, do you have any questions for me about the role or the company?",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-3070354033281820859.wav"
                    }
                },
                {
                    "time": "13:36:09",
                    "content": "**John**: [proper speed and proper volume] Like my my I want salary 4000 can.",
                    "metadata": {
```

```
                        "path": "files/johndoe/NPC-learner--
366238676448395143.wav",
                        "word_cnt": 8,
                        "duration": 3.48,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:36:38",
                    "content": "**Emily**: [understanding]  John, I
appreciate your honesty about your language
skills.[supportive]  It's clear that you have a strong team spirit
and a willingness to help others, which are valuable traits in our
company.[professional]  Regarding the salary, let's discuss the
offer further during the negotiation stage.[professional] We aim to
compensate our employees fairly based on their skills and
contributions.[encouraging]  I believe with your attitude and
potential, we can find a suitable arrangement that works for both
parties.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
8366907553940102034.wav"
                    }
                },
                {
                    "time": "13:36:51",
                    "content": "**John**: [fast and proper volume]
If no 4000 I don't want this job law like all my friend are all very
also also 4000.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
8395557179546118708.wav",
                        "word_cnt": 19,
                        "duration": 6.7,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:37:11",
```

```
                    "content": "**Emily**: [empathetic]  John, I
understand your concerns about the salary offer.[encouraging]  Let's
work together to find a middle ground that reflects both your skills
and the value you bring to our team.[supportive]  Your dedication
and willingness to learn are qualities we truly appreciate here at
ABC IT Company.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
3190124631195117613.wav"
                    }
                },
                {
                    "time": "13:37:19",
                    "content": "**John**: [proper speed and proper
volume] OK. So so 4000, OK.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
8263025383733173182.wav",
                        "word_cnt": 5,
                        "duration": 2.44,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:37:35",
                    "content": "**John**: [fast and proper volume]
So can can give me 4000 ringgit?",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
4552235756481026864.wav",
                        "word_cnt": 7,
                        "duration": 2.72,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "13:37:54",
```

```
                    "content": "**John**: [fast and soft] Uh, why? 
Why? Why you keep silent? Can you can your company give 4000 or 
not?",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
2593615291242980760.wav",
                        "word_cnt": 16,
                        "duration": 5.44,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "13:38:16",
                    "content": "**Emily**: [calm and 
reassuring]  John, I appreciate your enthusiasm and eagerness to 
discuss your salary expectations.[calm and reassuring] Let's take a 
moment to understand each other's perspectives.[calm and reassuring] 
It's important for us to find a fair and mutually beneficial 
agreement.[calm and reassuring] Let's discuss how we can align your 
expectations with the company's guidelines.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
8104271044927709962.wav"
                    }
                },
                {
                    "time": "13:38:29",
                    "content": "**John**: [fast and soft] OK, OK, so 
so can I I I won 4000 ringgit. Don't don't beat around the Bush, 
OK?",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
5910460283960459776.wav",
                        "word_cnt": 18,
                        "duration": 6.8,
                        "loudness": "soft"
                    }
                }
            ],
```

```
            "speech_rate": 139.3302371058421,
            "speech_loudness": "soft"
        }
```


Good Trainee with Bad Interviewer


```
"feedback": {
            "world_name": "Job Interview",
            "world_desc": "a job interview which involves in John
Doe, an interviewee with Computer Science background and Emily, a
demanding interviewer who do not have much patience. The interview
session takes place in meeting room, ABC IT company.",
            "overall_feedback": {
                "criterion 1": {
                    "name": "Expressing positive attitude",
                    "score": 2,
                    "reason": "Strengths: John showed some positive
attitude by trying to move the conversation forward and checking on
Emily's well-being. Weaknesses: John's attempts to keep the
conversation light-hearted were not very effective as Emily was
focused on getting concise answers. Improvement: John could show
more enthusiasm and positivity throughout the interview, even when
under pressure."
                },
                "criterion 2": {
                    "name": "Expressing confidence in a polite
tone",
                    "score": 3,
                    "reason": "Strengths: John maintained a polite
tone throughout the interview, even when Emily was impatient. He
spoke at a proper speed and volume most of the time. Weaknesses:
John's confidence seemed to waver slightly when under pressure to
provide concise answers. Improvement: John could work on maintaining
a more consistent level of confidence and assertiveness in his
responses, especially when asked to be more concise."
                }
            },
            "transcript": [
                {
                    "time": "14:19:35",
```

```
                    "content": "**Emily**: [impatient]  Time is
ticking, John.[impatient] Let's get straight to the
point.[impatient] Can you showcase your problem-solving skills in a
real-world scenario for me?[impatient] And remember, we need people
who can handle stress and work overtime.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
3966132902365027264.wav"
                    }
                },
                {
                    "time": "14:20:39",
                    "content": "**John**: [proper speed and soft] Hi
Emily, so my name is John and I come from Utah having the degree of
computer science. So I can share one of my most challenging projects
in my university studies which is involving in the large language
models development. So the leader and I try to guide my team to
develop the large language models specifically for the question and
answering and.In the end we reached the 90% of the accuracy. So
throughout the development we face a challenge which involving in
the resource constraint. So because of the GPU, if we were to use
the GPU with more RAM, it is actually a bit costly for us. So
eventually we apply the quantization technique and we resolve the
view constrain.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
3359579532332024333.wav",
                        "word_cnt": 124,
                        "duration": 56.84,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "14:20:55",
                    "content": "**Emily**: [frustrated]  John, I
appreciate your detailed explanation, but we need to keep it
concise.[frustrated] Can you give me a brief summary of how you
handled the resource constraint issue?[frustrated] We are running
out of time here.",
```

```
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
3935981205562115223.wav"
                    }
                },
                {
                    "time": "14:21:05",
                    "content": "**John**: [proper speed and proper
volume] Alright, so basically it is a sort of quantisation technique
that effectively reduce the memory consumption, yeah.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
6454636324065925736.wav",
                        "word_cnt": 17,
                        "duration": 7.28,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:21:21",
                    "content": "**Emily**: [impatient]  John, let's
move on to the next question.[impatient] Can you provide an example
of a project where you had to work under tight deadlines and how you
managed the stress?[impatient] And remember, keep it brief.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
5578524205655879058.wav"
                    }
                },
                {
                    "time": "14:21:50",
                    "content": "**John**: [fast and soft] Alright so
for I can use the same project as an example. So actually the
project is like I need. We only have two weeks of time to finish
everything and how I how we handle the stress is I try to split the
work evenly and let each of run the notebook screen concurrently and
in the end we managed to submit the thing on time.",
                    "metadata": {
```

```
                        "path": "files/johndoe/NPC-learner-
6494046802105733642.wav",
                        "word_cnt": 67,
                        "duration": 25.48,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "14:22:05",
                    "content": "**Emily**: [impatient] John, we are
almost out of time here.[impatient] Can you quickly explain how you
would approach a complex AI problem that requires innovative
solutions?[impatient] Remember, I need a concise answer.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
4118880763993062521.wav"
                    }
                },
                {
                    "time": "14:22:18",
                    "content": "**John**: [proper speed and proper
volume] Alright, to suffer an AI project we actually need to 1st
understand the problem and then try to gather the sufficient data
for the training.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
4763156478728168173.wav",
                        "word_cnt": 25,
                        "duration": 10.08,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:22:37",
                    "content": "**Emily**: [frustrated]  John, I
need you to be more concise and to the point.[frustrated] We don't
have much time left.[frustrated] Can you give me a specific example
of a complex AI problem you've solved in the past and how you
```

```
approached it?[frustrated] Remember, I need a clear and direct
response.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
6920373690020368400.wav"
                    }
                },
                {
                    "time": "14:22:59",
                    "content": "**John**: [fast and proper volume]
Alright, so the one of the AIA challenge is I do image segmentation
for the traffic sign purpose and it is actually challenging because
we don't have sufficient data set and how we eventually solve it is
we try to use unsupervised learning method and that's it.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
6991870736165699844.wav",
                        "word_cnt": 47,
                        "duration": 18.72,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:23:29",
                    "content": "**Emily**: [impatient]  John, we are
almost out of time here.[impatient]  Can you quickly explain how you
would approach a complex AI problem that requires innovative
solutions?[impatient]  Remember, I need a concise
answer.[frustrated]  John, I need you to be more concise and to the
point.[frustrated]  We don't have much time left.[exasperated]  Can
you give me a specific example of a complex AI problem you've solved
in the past and how you approached it?[exasperated]  Remember, I
need a clear and direct response.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
1105998557881174138.wav"
                    }
                },
                {
```

```
                    "time": "14:23:46",
                    "content": "**John**: [fast and proper volume]
Yeah, I think that summarize what I thought just now. So maybe can
we move on to the topic.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
4314179540436564466.wav",
                        "word_cnt": 19,
                        "duration": 6.68,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:23:59",
                    "content": "**John**: [fast and proper volume]
Are you OK?",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
5371181029116148420.wav",
                        "word_cnt": 3,
                        "duration": 0.96,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:24:09",
                    "content": "**John**: [fast and soft] I think
the time collection we can move on to the salary topic.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
9070666028107808702.wav",
                        "word_cnt": 13,
                        "duration": 3.16,
                        "loudness": "soft"
                    }
                },
                {
                    "time": "14:24:34",
```

```
                    "content": "**John**: [fast and proper volume] 
Alright, I'm sorry but I can roughly run out for my personal skills. 
So basically I have a problem solving skills and also debugging 
skills.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
2130630088717654868.wav",
                        "word_cnt": 25,
                        "duration": 8.64,
                        "loudness": "proper volume"
                    }
                }
            ],
            "speech_rate": 147.99767846778872,
            "speech_loudness": "proper volume"
        }
```


Good Trainee with Good Interviewer


```
"feedback": {
            "world_name": "Job Interview",
            "world_desc": "a job interview which involves in John 
Doe, an interviewee with Computer Science background and Emily, an 
inspiring interviewer who likes to talk with youngsters. The 
interview session takes place in meeting room, ABC IT company.",
            "overall_feedback": {
                "criterion 1": {
                    "name": "Positive Attitude",
                    "score": 4,
                    "reason": "Strengths:\n- John displayed a 
positive attitude throughout the interview, showing enthusiasm and 
excitement about his projects.\n- He was appreciative of the 
interviewer's questions and feedback.\nWeaknesses:\n- John could 
have shown more enthusiasm when discussing the technical aspects of 
his projects.\nImprovement:\n- Maintain consistent enthusiasm and 
positivity throughout all aspects of the interview."
                },
                "criterion 2": {
                    "name": "Confident Communication",
                    "score": 3,
```

```
                    "reason": "Strengths:\n- John communicated
confidently and clearly about his projects and technical skills.\n-
He maintained a polite tone throughout the
interview.\nWeaknesses:\n- At times, John spoke too fast, which
could have affected the clarity of his
explanations.\nImprovement:\n- Slow down the pace of speech to
ensure better clarity and understanding for the interviewer."
                }
            },
            "transcript": [
                {
                    "time": "14:09:18",
                    "content": "**Emily**: [friendly]  Hi
John![friendly] It's great to have you here today.[friendly] Why
don't you start by telling me a little bit about yourself and your
background in computer science?",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
6425961227708376530.wav"
                    }
                },
                {
                    "time": "14:10:00",
                    "content": "**John**: [proper speed and proper
volume] Hi I'm Lee so my name is John and I am graduated from Utah
studying Computer Science. So basically I can showcase some of my
problem solving skills to you. Last time I have an AI assignment
which is involving in training the large models. I as a team leader
to lead the whole theme to develop a large language models that can
be used for general question answering and as a result we achieve
like 90% of the accuracy.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
6681134005238909347.wav",
                        "word_cnt": 80,
                        "duration": 34.84,
                        "loudness": "proper volume"
                    }
                },
```

```
                {
                    "time": "14:10:18",
                    "content": "**Emily**: [impressed]  Wow, John,
that sounds like a challenging and rewarding project you
led![curious]  Can you share more about the specific challenges you
faced while developing the large language model?[curious] How did
you and your team overcome them to achieve such high accuracy?",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
60328958683180790.wav"
                    }
                },
                {
                    "time": "14:11:08",
                    "content": "**John**: [fast and proper volume]
Alright, so one of the main challenge in the project is we have
limited resource constraints. For example, if we are to use a better
GPU with more RAM, we need to subscribe for the Google Collab Pro.
But it is a bit costly for us. So how we solve it is to we learn the
quantization techniques which we can effectively reduce the memory
consumption for the GPU. So eventually we can train the model
using.GPU T4 which is also provided by Google Collette and it is
free.And how we eventually reached a currency is because we select a
good enough baseline model for the fine tuning. And that explains
everything I guess.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
3361042071068128167.wav",
                        "word_cnt": 116,
                        "duration": 44.88,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:11:30",
                    "content": "**Emily**: [impressed]  John, your
resourcefulness in overcoming those challenges is truly
commendable.[impressed] It's clear that you and your team put a lot
of thought and effort into finding creative
```

```
solutions.[curious]  Could you elaborate on how the quantization
techniques specifically helped in reducing memory
consumption?[curious] I'm intrigued to learn more about the
technical aspects of your approach.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
15627523893769740.wav"
                    }
                },
                {
                    "time": "14:12:12",
                    "content": "**John**: [proper speed and proper
volume] Alright, so the main quantization technique is basically is
trying to shrink the floating numbers from like let's say 64 bits is
double and we string it into 32 and then 16 and for. Actually there
are a lot of this sort of numerical computing is actually
dynamically designed, which is the right precision. So like the
lowest we can reach until floating .16 and for some we just remain
the floating 32. So it actually effectively cut the memory
concentration by.By half and by quarter.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
8610368722395622842.wav",
                        "word_cnt": 87,
                        "duration": 35.76,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:12:31",
                    "content": "**Emily**: [impressed]  John, your
understanding and application of quantization techniques is truly
remarkable.[curious]  I'm curious to know, did you encounter any
challenges or obstacles while implementing this
approach?[encouraging]  It's important to learn from both successes
and setbacks in order to grow as a professional.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
6523388032729457638.wav"
```

```
                    }
                },
                {
                    "time": "14:12:57",
                    "content": "**John**: [fast and proper volume]
Alright, so basically the main challenge is of course we need to
integrate with some third party libraries, so the integration is a
bit challenging and of course in the end as we realize that the
setting can be.Referenced by the Hanging Face channel, and I think
that's all about the project.",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner-
7689817765256850727.wav",
                        "word_cnt": 53,
                        "duration": 20.0,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:13:16",
                    "content": "**Emily**: [curious]  That sounds
like quite the project, John.[curious] Integrating with third party
libraries can definitely pose some challenges.[encouraging]  It's
great to see how you navigated through it.[encouraging] Can you
share more about the impact your project had within the team or
organization?",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural--
8634052197175088673.wav"
                    }
                },
                {
                    "time": "14:13:40",
                    "content": "**John**: [proper speed and proper
volume] Alright, so basically the main impact is this can be used
for commercial purpose, like the end users can actually use it for
some Q&A. And I think I cannot share too much because there are some
confidential insight. But yeah, that's all I can share to you.",
                    "metadata": {
```

```
                        "path": "files/johndoe/NPC-learner-
2489882398854222390.wav",
                        "word_cnt": 48,
                        "duration": 19.2,
                        "loudness": "proper volume"
                    }
                },
                {
                    "time": "14:13:57",
                    "content": "**Emily**: [enthusiastic]  That's
fantastic to hear, John![enthusiastic] Your project's potential for
commercial use is truly impressive.[inquisitive]  I completely
understand the need for confidentiality, but could you perhaps share
a bit more about the technical aspects of how the Q&A system was
implemented?[supportive]  Feel free to share as much as you're
comfortable with, we're here to learn from your experiences.",
                    "metadata": {
                        "path": "files/johndoe/NPC-JennyNeural-
7850922598586200756.wav"
                    }
                },
                {
                    "time": "14:14:20",
                    "content": "**John**: [fast and proper volume]
Yeah, so basically beside our problem solving skills, I also know
how to debug a program very well. So I know a variety of languages
like C++, Python, Java, C# and I think that the time constraint is
here, so maybe we should move forward still the salary, shall we?",
                    "metadata": {
                        "path": "files/johndoe/NPC-learner--
5303796903710556O.wav",
                        "word_cnt": 50,
                        "duration": 18.72,
                        "loudness": "proper volume"
                    }
                }
            ],
            "speech_rate": 150.17301038062286,
            "speech_loudness": "proper volume"        }
```

Appendix C – Weekly Log

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 1**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE

Improving Project 1 Coding and resolves a few limitations, including: avatar forgets their role, long waiting which mainly due to bad architecture design. Applied OOP concepts to separate different threads like Brain, Sensor so can scale to better multimodality in future.

Tested smaller and open-source model like LLaMa 7B and Mistral 7B with competitive performance, shrink the waiting time by constraining the output token.

2. WORK TO BE DONE

Use rule-based method to resolve the limitation where LLM prefers to “speak” rather than “stay silent”.

3. PROBLEMS ENCOUNTERED

Avatars fail to know that others are talking due to the characteristics of “event-driven” simulation than “time-driven” to save computational cost. LLM is computational expensive, inspired by the concept of Mixture of Expert, should resolve this by designing smaller algorithm/ rules.

4. SELF EVALUATION OF THE PROGRESS

On track and everything should work great.

Supervisor’s signature

Student’s signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	Study week no.: 2
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE

Defined multiple scenario (1) Family Conflict (2) Couple Conflict (3) Sales at xxx (4). Handled the way agents perceive interruption event.

2. WORK TO BE DONE

Discuss with soft skills lecturer to design (user-centric design) multiple scenarios for testing and validation.

Define required user interface.

3. PROBLEMS ENCOUNTERED

Speak interruption may need some small machine learning model.

4. SELF EVALUATION OF THE PROGRESS

Good progress.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 3**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Refactored coding work for both Unity and python server.
2. WORK TO BE DONE Define specific criteria for system evaluation.
3. PROBLEMS ENCOUNTERED Avatar response is slow and might have impact the user experience.
4. SELF EVALUATION OF THE PROGRESS The project is progressing smoothly.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 4**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Explored multiagent capability for more natural conversation.
2. WORK TO BE DONE Seek a probabilistic model to determine whether is continue listen or speak
3. PROBLEMS ENCOUNTERED Validation of the model accuracy can be challenging.
4. SELF EVALUATION OF THE PROGRESS The project is progressing steadily.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 5**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Drop unneeded features for simplicity.
2. WORK TO BE DONE Make the evaluation UI for trainers.
3. PROBLEMS ENCOUNTERED Project scope is a bit large and should narrow further.
4. SELF EVALUATION OF THE PROGRESS Needs to ensure all the features can be evaluated properly.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 6**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Done UI for trainers.
2. WORK TO BE DONE Refine the coding work for system stability.
3. PROBLEMS ENCOUNTERED No so far.
4. SELF EVALUATION OF THE PROGRESS The project is progressing steadily.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 7**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Done project demonstration to soft skills lecturers.
2. WORK TO BE DONE Refine testing examples according to soft skills lecturers' suggestions.
3. PROBLEMS ENCOUNTERED A need of better demonstration.
4. SELF EVALUATION OF THE PROGRESS The progress is still on track.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 8**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Performed testing against different scenarios defined.
2. WORK TO BE DONE Slightly adjust the program to meet the testing criteria.
3. PROBLEMS ENCOUNTERED Avatars response time needs to be improved.
4. SELF EVALUATION OF THE PROGRESS Real time issue is a serious concern and needs to be resolved as soon as possible.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 9**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Adjusted LLM modules and performed system testing.
2. WORK TO BE DONE Perform testing on time efficiency.
3. PROBLEMS ENCOUNTERED Real time speed is stunted by python asyncio module.
4. SELF EVALUATION OF THE PROGRESS Needs to think out of the box to solve the issue.

Supervisor's signature

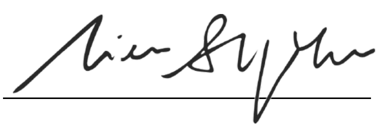
Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 10**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE
Second meeting with soft skills lecturers.
2. WORK TO BE DONE
Run testing for different avatars with different trainees.
3. PROBLEMS ENCOUNTERED
User experience issues which is the VR avatars are not so realistic.
4. SELF EVALUATION OF THE PROGRESS
Progress is still fine.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Year 3 Trimester 3	**Study week no.: 11**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE Rolling out two-phase simulation module that resolves system delay and inefficiencies.
2. WORK TO BE DONE Updating the source code and perform final evaluation.
3. PROBLEMS ENCOUNTERED No.
4. SELF EVALUATION OF THE PROGRESS Progress is still acceptable.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

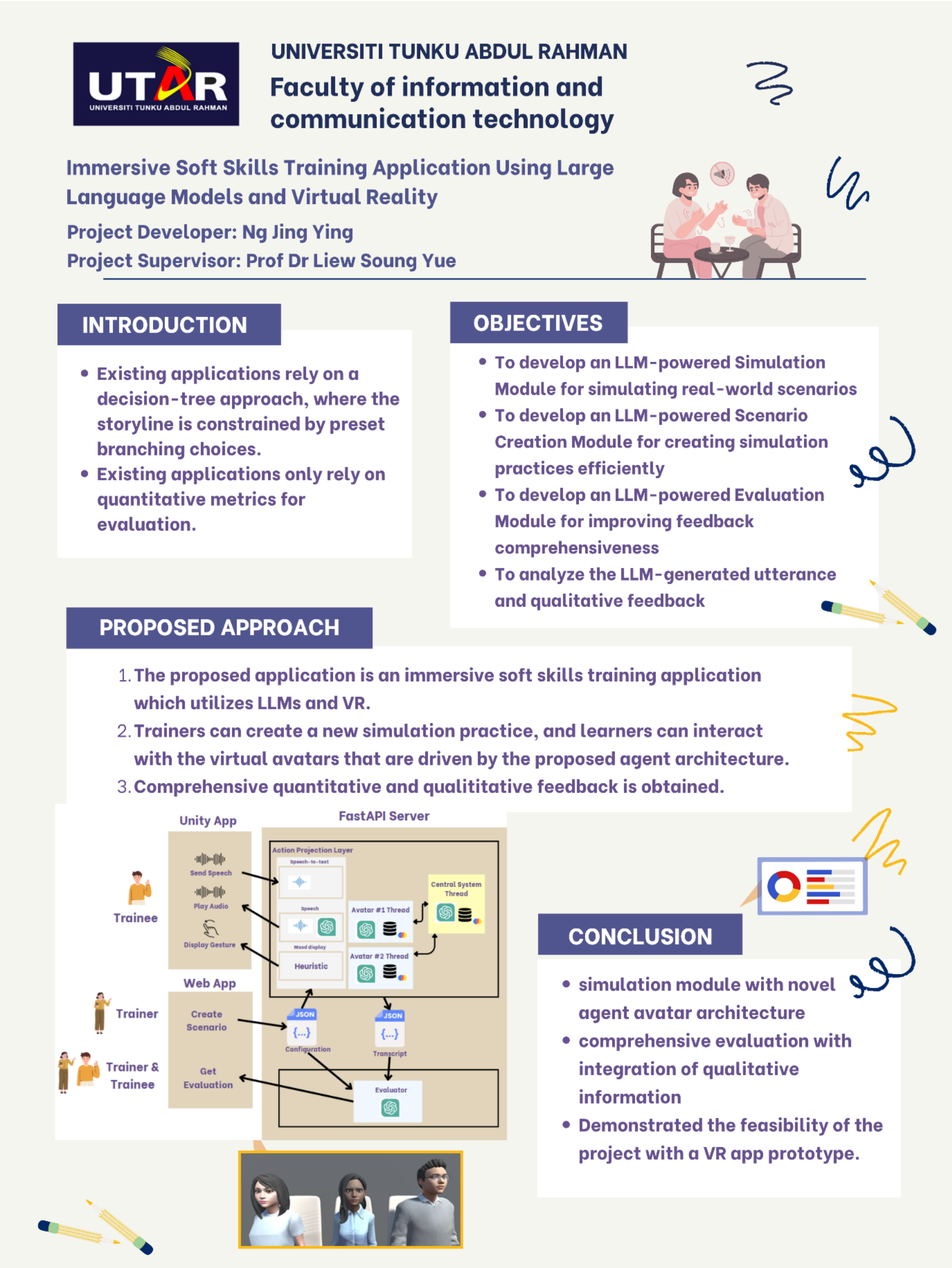
Trimester, Year: Year 3 Trimester 3	**Study week no.: 12**
Student Name & ID: Ng Jing Ying 21ACB01845	
Supervisor: Prof Ts Dr Liew Soung Yue	
Project Title: IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY	

1. WORK DONE System testing completed. Project objective reached.
2. WORK TO BE DONE Finalizing report and presentation.
3. PROBLEMS ENCOUNTERED No.
4. SELF EVALUATION OF THE PROGRESS Project is completed.

Supervisor's signature

Student's signature

Appendix D - Poster



Appendix E – Plagiarism Check Result

Document Viewer

Turnitin Originality Report

Processed on: 25-Apr-2024 23:47 +08
ID: 2360920858
Word Count: 15384
Submitted: 3

21ACB01845_FYP2_turnit.docx By Jing Ying Ng

Similarity Index	Similarity by Source	
3%	Internet Sources:	3%
	Publications:	1%
	Student Papers:	2%

exclude quoted | include bibliography | excluding matches < 8 words | mode: quickview (classic) report | print | download

1% match (Internet from 10-Oct-2013)
http://www.epubbud.com

<1% match (Internet from 10-Oct-2022)
http://eprints.utar.edu.my

<1% match (Internet from 13-Mar-2024)
http://eprints.utar.edu.my

<1% match (Internet from 20-Jan-2024)
http://eprints.utar.edu.my

<1% match (Internet from 15-Dec-2022)
http://eprints.utar.edu.my

<1% match (Internet from 10-Oct-2023)
http://eprints.utar.edu.my

<1% match (Internet from 09-Oct-2022)
http://eprints.utar.edu.my

<1% match (Internet from 15-Dec-2022)
http://eprints.utar.edu.my

<1% match (Internet from 03-Mar-2023)
http://eprints.utar.edu.my

APPENDIX

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Ng Jing Ying
ID Number(s)	21ACB01845
Programme / Course	Bachelor of Computer Science
Title of Final Year Project	**IMMERSIVE SOFT SKILLS TRAINING APPLICATION USING LARGE LANGUAGE MODELS AND VIRTUAL REALITY**

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 3 % **Similarity by source** Internet Sources: 3 % Publications: 1 % Student Papers: 2 %	Within the required range.
Number of individual sources listed of more than 3% similarity: 0	Within the required range.
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: Liew Soung Yue

Date: 26/4/2024

Signature of Co-Supervisor

Name: ______________________

Date: ______________________

FYP2 Checklist

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student ID	21ACB01845
Student Name	Ng Jing Ying
Supervisor Name	Prof Ts Dr Liew Soung Yue
TICK (√)	**DOCUMENT ITEMS** Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
N/A	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

____

(Signature of Student)

Date: 25/4/2024