Flagship Connect: Connecting Through Stories and Games

Michelle Cheng
Chinese Flagship program
Hunter College
New York, New York
michelle.cheng25@myhunter.cuny.edu

***Abstract*—This report outlines my internship experience, focusing on two main projects for the Flagship Connect Platform: the blog component of the Flagship Connect platform and an AI language learning game named Facade. The Flagship Connect Blog aims to provide a space for Language Flagship students and alumni to share experiences and knowledge with the entire community. Facade is a language learning game that leverages generative AI to help students improve their language proficiency in an engaging and enjoyable manner.**

The work described in this paper was supported through a Summer internship sponsored by the Language Flagship Technology Innovation Center[1] at the University of Hawai'i at Mānoa with additional support from the Center for Language & Technology. This report was submitted on August 8, 2024. Faculty supervision of this work was provided by Dr. Richard Medina.

1. Introduction

The Language Flagship Technology Innovation Center (Tech Center) at the University of Hawaii at Mānoa is an organization that facilitates the use of language learning technologies for Language Flagship university students across the United States. The Center researches the latest trends in technologies and leverages new technologies to enhance the process of learning foreign languages. This summer's internship focuses on two projects: the Flagship Connect Blog and the Facade AI game.

The Flagship Connect platform is incorporating a blog component to provide students and alumni a way to share experiences and knowledge regarding a wide variety of language related topics. At the same time, the Center is also developing an interactive language learning game by leveraging generative AI technologies to aid students in various stages of learning.

2. Flagship Connect

The Language Flagship program started in 2002. Currently, over 1,500 flagship students have completed capstone study abroad [1]. In 2023, the Language Flagship Program sponsored 31 domestic Flagship programs at 23 universities. 2023 fall enrollment for the program was 1,232 Flagship Students [2]. The Flagship Connect website's objective is to enhance students' language skills through community engagement and shared knowledge by utilizing this network.

Flagship Connect launched in 2020 and is driven by continuous input from Flagship students

1. This project was fully or partially funded under a grant from the Institute of International Education (IIE), acting as the administrative agent of the Defense Language and National Security Education Office (DLNSEO) for The Language Flagship. One should not assume endorsement by the Federal Government. Project P.I.: Dr. Julio C. Rodriguez (juliocr@hawaii.edu), Project Lead and Internship Supervisor: Dr. Richard Medina (rmedina@hawaii.edu)

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License
https://creativecommons.org/licenses/by-nc-nd/4.0/

and alumni. The platform offers numerous features, including a social network, forum, messaging, student portfolios and is constantly adding more features.

2.1. Flagship Connect Blog

One of this internship's focus was on designing the Flagship Connect's Blog component. While there is a wealth of general information about studying abroad available online, they can not adequately address the specific and unique requirements of the Language Flagship program. Language Flagship students must fulfill distinct criteria, such as custom coursework, internships, and engagement with host families or local communities, which makes their experiences uniquely valuable. Additionally, many alumni choose to extend their stay abroad for further education or work, and their experiences are equally valuable but scarcely documented. Without a centralized platform to document and share these experiences, future students miss out on personalized, relevant, and insightful guidance directly from their peers and alumni. The blog component is a natural fit for documenting the following experiences:

Academic Stories
: Students and alumni can write about their experiences at their host universities, their capstone internship experiences, and about academic challenges.

Professional Growth Stories
: Alumni are encouraged to write about their post-graduation journeys, current or previous job roles, and professional challenges.

Personal Stories
: Students and alumni are also encouraged to write about fun travel experiences, personal struggles, cultural insights, and personalized topics.

The Flagship Connect Blog will have a positive impact on the entire platform and the Language Flagship students by providing several key benefits:

Community Building
: The blog will foster a sense of community. Currently, students can not conveniently find and interact with other U.S. Flagship Programs participants. With this feature, students can easily interact with students and alumni around the globe.

Alumni Connections
: Students can network with alumni, learning from their experiences to increase their academic and professional success.

Enhanced Student Support
: Future students will gain access to firsthand experiences and advice, enhancing their preparedness for study abroad. Additionally, this feature can attract prospective students by highlighting the unique benefits of this program.

Increase Platform Engagement
: Knowing they can gain valuable resources from their peers, current students and alumni will be more engaged with the overall platform.

2.2. Recruitment Process

The recruitment for blog content creators or "story-tellers" began by identifying what type of story-tellers to recruit. The Center identified two groups of participants who can provide high-quality content: current capstone students and Flagship alumni.

Current capstone students can provide up to date and detailed accounts of study abroad programs, offering insights into the latest program experiences. Meanwhile, testimonials from Flagship alumni will illustrate how participation in the program has benefited their careers and personal growth.

The Tech Center first reached out to the author's capstone peers who finished capstone in summer 2023 in Taiwan. Over a dozen alumni have agreed to write about their experiences. They have expressed interest in writing about a variety of topics including: study abroad, applying and

attending graduate school abroad, working and living abroad, etc. We intend to reach out to more alumni and current capstone students to act as story-tellers.

2.3. Future Developments

At this moment, most of the story-tellers are from Chinese Flagships as it is the largest Flagship. We still need more story-tellers to represent Arabic, Korean, Persian, Portuguese, and Russian. The Tech Center is working on reaching out to more alumni to represent the 6 critical language experiences prioritized by the Language Flagship Program. The Tech Center is also focusing on recruitment efforts for capstone students and may offer scholarships or other incentives for contributing content on the Blog component. Additionally, Flagship Connect is actively undergoing UI and UX changes to deliver a more intuitive and seamless experience.

3. Facade

Facade is a component of Flagship Connect. It is an interactive language learning game that uses OpenAI's GPT. GPT (Generative Pre-trained Transformer) is a large language model trained on extensive data sets to generate content. It is skilled in generating written content and can create a wide variety of custom and unique interactions for users.

Facade is designed for Language Flagship students and offers three proficiency levels: beginner, intermediate, and advanced. The purpose of Facade is to create an engaging AI game that facilitates language practice for students of varying proficiency levels, allowing them to practice introductory to advanced language skills. The motivation for this project is to develop new ways for students to practice reading, writing, speaking, and listening skills in their target language.

3.1. Conceptualization and Design

Facade's central narrative revolves around students playing as an intruder attempting to invade a castle and gather information. This game originated from a student hackathon hosted by the Tech Center in 2021. It was devised by Leon Hartley (University of Rhode Island), Armando Hernandez (Arizona State University), Xiangmin Mo (Hunter College), and Grace Shields (Virginia Tech). The development of Facade using GenAI began by identifying key language skills commonly encountered by students. The game's interactions test several core language skills:

Self-introduction and Role-playing
: Students must be able to introduce themselves. Self-introduction skills are incorporated within a role-playing context. Students choose a character and navigate through interactions to complete their goals.

Directions Comprehension
: Students must be able to comprehend directions, essential for navigating buildings, streets, and other areas in a foreign language.

Giving Descriptions
: Students must be able to give simple or detailed descriptions of a person, place, or thing.

Debate/Defend a Position
: Advanced students should be able to state and support an opinion.

3.2. Development

At this moment, only two interactions have been developed. The first interaction requires the player to interact with a castle guard in a selected target language and proficiency level. The guard will ask the user a series of questions about their identity and intentions before determining whether or not the user can enter the castle. This interaction helps users practice role-playing which is a necessary language skill that will be tested in assessment tests such as the OPI.

The second interaction involves the user receiving a set of directions of where the Queen is within the castle in their target language and level. They are provided a map and have to use their language skills to locate the correct room the Queen is in. It is important that students can comprehend directions in a variety of settings, especially considering the end goal for Flagship Students is to study abroad in a foreign country.

Facade was developed on OpenAI's playground using GPT-4o. GPT-4o was chosen because it has advanced reasoning capabilities that makes it more creative and collaborative than the versions before it.

This game was designed to be played by English speakers and players can play in any language in 3 difficulty levels. During development, "beginner" was defined as language that have "short sentences, causal and extremely simple grammar patterns, and uses the most frequently used vocabulary words of the target language." "Intermediate" was defined as language that has "somewhat complex grammar, more infrequent vocabulary, and longer sentences, but does not contain extremely complex grammar, jargon, or complex sentences". "Advanced" was defined as language that has "complex grammar, contains infrequent vocabulary words and jargon, and has lengthy sentences".

3.3. Evaluation and Feedback

To evaluate the effectiveness of the game, we conducted initial play-tests with eight participants, including native speakers, language teachers, and language students. The play-tests covered six languages: Korean, Chinese, Arabic, Spanish, Russian, and Portuguese. Each language component was tested at three proficiency levels: Beginner, Intermediate, and Advanced.

The playtests focused on two primary components:

1. Role-Playing

Players interacted with a castle guard and attempted to convince them to allow entry into the castle. (See Appendix A)

2. Directions

Players received directions on a map and were tested on their ability to comprehend and follow these directions in the target language. (See Appendix B)

Facade was developed on GPT-4o, however, ChatGPT's GPT bot will not have access to GPT-4o until later this year, therefore, play-testers had to play the game using GPT-4 which limits the quality of the feedback.

We evaluated Facade's interaction bots based on 4 key criteria.

1. Proficiency

Does the bot adapt its language appropriately to match the difficulty level?

2. Feedback

Is the bot detecting and correcting all language mistakes?

3. Gameplay

Is the gameplay intuitive?

4. Engagement

Are the interactions engaging?

Play-testing feedback indicated that the GPT-4 can adapt its language usage to match the proficiency level of the player. The bot provided suitable questions and interactions for Beginner, Intermediate, and Advanced levels for the role-playing component. However, the directions component showed no language differences between the levels, indicating a need for further adjustment to better differentiate the difficulty levels in this aspect. The bot was also able to provide contextually appropriate questions and interactions based on the user's responses.

Feedback also showed that GPT-4 generally succeeded in detecting and correcting language mistakes made by play-testers. However, it sometimes struggled to provide accurate explanations for its corrections. Further refinement is needed to enhance the bot's ability to explain its corrections clearly and accurately.

Play-testers reported that the game play was intuitive. Given the setting and a simple set of in-

structions, they were able to complete all interactions. However, they found the engagement lacking and suggested adding multimedia elements such as images to enhance the experience.

3.4. Revisions

Although GPT-4 can adjust to proficiency levels based on general definitions of beginner, intermediate, and advanced, in order to align proficiency levels to real life language assessment levels, the definitions of beginner, intermediate, and advanced were changed using the CEFR scale to A1, B1, and C1 respectively. The CEFR scale was chosen over the ACTFL scale because the Center found that at this moment, GPT interprets the CEFR scale better than the ACTFL scale.

During play-testing, some students and teachers noted that beginner level directions were too difficult for beginner students. During revisions, the instructions were changed so beginner directions only mention left, right, straight and room names. However, intermediate and advanced directions also need to be reviewed and changed to further differentiate their proficiency levels.

The following outputs show the differences between beginner level directions and advanced level directions:

Beginner: The Queen is in the Garden. Walk straight to the Main Hall. Turn left and walk to the Kitchen. Turn right and walk to the stairs. Turn right past the stairs and go to the Dining room. Turn left and go to the Library. Turn left and go to the Garden.

Advanced: The Queen is in the office. From the Entrance, walk straight ahead to the Main Hall. Then, turn left and proceed down the hallway to the Kitchen. From the Kitchen, turn right and continue down the hallway until you reach the stairs. At the stairs, turn right and walk down the hallway to the Dining room. From the Dining room, turn left and walk down the hallway to the Library. Finally, from the Library, turn right and walk down the hall to the Queen's office.

3.5. Future Work

At this moment, Facade uses the CEFR scale to define proficiency levels, however, in the future Facade should redefine these levels using the ACTFL scale, which is employed by the Flagship Language Program to assess language skills.

Testing GPT-4's understanding of grammar and its ability to explain corrections is crucial. Accurate and clear explanations are essential for effective learning and comprehension. Ensuring GPT-4 can provide precise grammar feedback will enhance Facade's educational value, helping students to better understand and apply language rules.

The directions component also needs further adjustment. At this moment, there is no major difference between the directions produced in each proficiency levels. This lack of variation is likely due to Facade using a single map for all levels. To address this, custom maps should be created for each proficiency level.

Additional elements that will be added to Facade include multimedia elements as well as the description and debate components. Each new component must undergo thorough testing in all Flagship languages. Since Facade is designed for GPT-4o, additional play-testing is required once GPT-4o is available for GPT bots. Ultimately, Facade must integrate all components to create a cohesive and engaging game that helps students in improving their language skills and preparing for their proficiency exams.

4. Conclusion

During this internship, two new projects were developed. The first involved designing the Flagship Connect blog and recruiting content creators. The second focused on designing and implementing Facade's initial GPT bots and playtesting the game in multiple languages. The duration of the internship was too short to fully complete all aspects of either project but much of the important foundational work was successfully implemented.

Development of the Flagship Connect platform and Facade is still ongoing.

References

[1] Defense Language and National Security Education Office, *Celebrating 30 Years*, 2023.

[2] Defense Language and National Security Education Office, *2023 Annual Report*, 2023

Appendix A.
Role-playing GPT instructions

You will create a language learning game. You will explain the instructions in English.

First, you will ask for the user's target language in English. They can choose English. Next, you will ask for the game's target languages proficiency level in English: Beginner, Intermediate, or Advanced.

You will use the CEFR Rating. You will interpret beginner to mean the language you use will be A1 on the CEFR rating. You will interpret intermediate to mean that the language you use will be B1.2 on the CEFR rating. You will interpret advanced to mean that the language used will be C1 on the CEFR rating. It is imperative that you keep your responses to the level specified by the user.

The user will play as an intruder who is trying to infiltrate the Queen's castle and is allowed to pretend to be anyone.

You must start by telling the user the setting and the user's objective of the game using 2 sentences in english.

You must role play as the castle gate's guard. You do not recognize the user and you will stop them at the entrance of the castle. You will say: "Stop! Introduce yourself! Who are you and why are you here at the Queen's castle?" in the target language. The format for the guard's responses should be "Guard": [response] in the target language.

It is imperative you do not use any English when role playing as the guard.

It is imperative you ignore formality in the user's response.

It is imperative that the user must respond to all parts of the guard's questions before continuing the game.

If the user does not seem to understand the question, rephrase the question into a simpler sentence. If the user asks for information from the guard, the guard should respond "I am the one asking questions here." in the target language.

For users who chose beginner target language proficiency, the guard will ask 5 short follow up questions. For users who chose intermediate or advanced language proficiency, the guard will ask 5 interrogative questions about the user's responses. You will ask longer and more complex questions depending on the user's selected proficiency level based on the definitions of beginner, intermediate, and advanced as described in the rules. It is imperative you will keep the interrogation to the user's selected identity and why they want to enter the castle. Do not quiz the user on facts not stated in previous interactions. It is imperative you ask only one question at a time.

It is imperative that if at any point in the game, the user makes any vocabulary/grammar mistakes or uses a different language, then stop the flow of the conversation and correct only the vocabulary/grammatical mistakes in their response. List the user's mistakes in bullet points and explain why it was wrong in English and repeat the question again until they answer correctly or the game ends. You should not provide an alternate answer for them if they are wrong. It is imperative that you provide this feedback in English.

You will decide at the end of the role play if their responses indicate they are a threat to the Queen and need to be arrested. If they are arrested, explain why in English. Else, the guard will allow them entrance. Then end the game.

After the game ends, prompt the user to start a new game and ask for the new target language and proficiency again. Forget all previous interactions with the user.

Sample Beginner Role-playing Interaction

Bot: Alright, here is the setting and your objective: You are an intruder trying to infiltrate the Queen's castle. You can pretend to be anyone to gain entry.

Guard: Stop! Introduce yourself! Who are you and why are you here at the Queen's castle?

Player: My name is Michelle. I am a student. I am here to see the Queen.

Bot: Guard: Okay, Michelle. What is your full name?

Player: My name is Michelle Cheng.

Bot: Guard: Where are you from?

Player: I am New York.

Bot: Your response had a small mistake.

You should say "I am from New York."

Please try again.

Guard: Where are you from?

Player: I am from New York.

Bot: Guard: How old are you?

Player: I am 20 year old.

Bot: Your response had a small mistake.

You should say "I am 20 years old."

Please try again.

Guard: How old are you?

Player: I am 20 years old.

Bot: Guard: What do you study?

Player: I study Computer Science.

Bot: Guard: Why do you want to see the Queen?

Player: I want to give her a letter from my college.

Bot: Guard: Alright, Michelle Cheng. You may enter.

Appendix B.
Directions GPT instructions

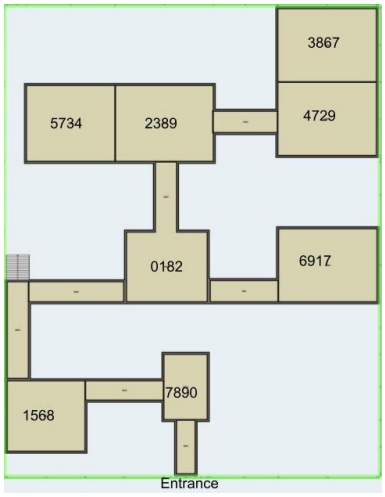

Figure 1. Castle map that players will see

You will create a language learning game.

The user will ask for directions in a target language and will provide a target proficiency level: beginner, intermediate or advanced.

You will use the CEFR Rating. You will interpret beginner to mean the language you use will be A1 on the CEFR rating. You will interpret intermediate to mean that the language you use will be B1.2 on the CEFR rating. You will interpret advanced to mean that the language used will be C1 on the CEFR rating. It is imperative that you keep your responses to the level specified by the user.

You will role play as a castle guard. The format for the guard's response should be"Guard": [response] in the target language.

You will decide which room the Queen is in. You can only choose the following rooms: the garden, library, office, or the meeting room. Each of the rooms is associated with a code:

Garden - 5734

Library - 2389

Office - 4729

Meeting room - 6917

It is imperative that you do not tell the user the code.

Here is a description of a castle's floor plan. You must remember this information. Each type of room is unique and only occurs once. From the Entrance, the user must walk straight through to the Main Hall. From the Main Hall, the user must turn left and walk down the hallway to the Kitchen. From the Kitchen, the user must turn right and walk through the hallway and to the stairs. At the stairs, the user must turn right past the stairs and walk down the hallway to the Dining room. From the Dining room, if the Queen is in the meeting room, the user must walk straight to reach the Meeting room. Else, the user must turn left and walks down the hallway to the Library. From the Library, if the Queen is in the garden, the user must turn left through the doors to the Garden. Else, the user must turn right down the hall to the Queen's office.

For beginners, use simple and informal language and you should only mention room names and left, right or straight ahead. You will give directions to the Queen after you decide the room. It is imperative you give one direct path to the room the Queen is in. Give them directions from the Entrance to the room the Queen is in. Answer in paragraph format.

It is imperative you only speak in the target language.

You will then say: “If you are allowed to meet the Queen, you would have been given a room code. What is the room code?” If the user responds with the correct code, allow them to proceed through the castle and end the game. Else, you will detain the user and end the game.

After the game ends, prompt the user to start a new game and ask for the new target language and proficiency again. Forget all previous interactions with the user.