

LingoComics: Situated Language Learning through Comic Style AI-generated Stories



Undergraduate Honours Thesis
Computer Science, Faculty of Science
Ontario Tech University

Deval Panchal

Supervisors: Mariana Shimabukuro & Dr. Christopher Collins

April 19, 2024

Abstract

There are many language learning tools available, and most of them are quite successful in their own purposes, as each offers a different learning approach. However, most of these apps focus primarily on memorization and vocabulary learning, and they lack the contextual aspect that highlights the distinction between learning a language. This thesis presents LingoComics, a situated language learning web application that uses comic-style AI-generated stories to provide a new way to learn languages. This approach leverages emotional engagement with content to enhance retention and practical language skills. LingoComics uses a simple framework of SvelteKit, Firebase, and OpenAI's API, a web application that would be accessible to a wider audience. By offering an innovative method of contextual learning, LingoComics aims to bridge the gap between standard vocabulary acquisition and practical language usage, potentially transforming how languages are learnt.

Table of Contents

| | |
|---------------------------------------------------------------------------|-----------|
| Abstract | 2 |
| 1 – Introduction | 4 |
| 1.1 – Motivation | 4 |
| 1.2 – Target Audience | 5 |
| 2– Background | 6 |
| 2.1 - LingoLand: Immersive Learning | 6 |
| 2.1.1 Personalized Immersion | 7 |
| 2.1.2 Generative World | 7 |
| 2.1.3 Encouraging Environment | 7 |
| 2.1.4 Conclusion | 8 |
| 2.2 Storyfier - Learning Comprehension | 8 |
| 2.2.1 Reading generated stories while covering target words | 8 |
| 2.2.2 Cloze test | 9 |
| 2.2.3 Use target words to write a story with adaptive AI assistance | 9 |
| 2.2.4 Conclusion | 10 |
| 2.3 - Exploring the Benefits and Challenges of AI-Language Learning Tools | 10 |
| 2.3.1 - Benefits of AI-Language Learning Tools | 10 |
| 2.3.2 - Challenges of AI-Language Learning Tools | 11 |
| 2.3.3 - Conclusion | 11 |
| 3. LingoComics | 12 |
| 3.1 - System Architecture | 13 |
| 3.1.1 - Wireframe | 14 |
| 3.1.2 - Tools | 16 |
| 3.1.2.1 - Sveltekit | 16 |
| 3.1.2.2 - Firebase | 17 |
| 3.1.2.3 - Google Cloud Service (GCS) | 17 |
| 3.1.2.4 - OpenAI API | 17 |
| 3.2 - Possible use case | 17 |
| 3.3 - Story designer | 18 |
| 3.3.1 - A comprehensive overview of contextual story generation | 19 |
| 3.3.2 - Comic style image generation | 21 |
| 3.3.3 - Challenges of creating a coherent story | 24 |
| 3.3.3.1 - Consistency in image generation | 24 |
| 3.3.3.2 - API limitations | 26 |
| 3.4 - Interactive activities | 27 |
| 3.4.1 - Drag-and-drop | 27 |
| 3.4.2 - Create your own Adventure | 29 |
| 4. Evaluation | 30 |

| | |
|-----------------------|-----------|
| 5. Future work | 31 |
| 5.1 - Achievements | 31 |
| 5.2 - Translations | 32 |
| 6. Conclusion | 33 |

1 - Introduction

Recent advancements in artificial intelligence have significantly impacted the educational and language learning experiences of individuals. Yet contextualization remains a challenging issue, especially with language learning applications. This thesis presents LingoComics. LingoComics is a web application that embeds AI-generated stories including both narrative and comic-style illustrations aiming to enhance language learning, in terms of contextualization and personalization. LingoComics offers relatable and situational narratives that are tailored to the learners' interest by applying an engaging comic-style illustration. LingoComics is designed to give learners creative control to generate contextually relatable stories using a simple and easy-to-use user interface to help them learn new vocabulary and practice contextual language use when creating stories or engaging in story-based activities.

This thesis presents related work (Chapter 2), prototype design, architecture, and implemented features (Chapter 3), future work and system evaluation plan (Chapter 4), and the final remarks and takeaways from the development of this project (Chapter 5).

1.1 - Motivation

Fortunately, there are many language learning tools available via downloadable apps or websites. Most language learning tools focus on memorization, grammar, and vocabulary building through repetition and personalized learning models, for instance, Duolingo and Memrise both focus on vocabulary learning which follows their own curriculum (Duolingo, n.d, Memrise, n.d). However, the existing language learning apps lack (1) curriculum flexibility – the ability to tailor your own lessons or curriculum to your preferred interests, and (2) language application or immersion such as generating appropriate language for different real-world contexts or situations, making target language practice toward fluency and vocabulary retention difficult.

Having an emotional connection with what you are learning helps contextualize concepts and improve retention. According to the study “How Does Fiction Reading Influence Empathy?” by Bal (2013), “stories generate empathy and emotional connection.” The ability to

author stories and control the content of lessons is often lacking in mainstream traditional language learning applications.

Thus, I designed and implemented LingoComics, to fill in the gaps related to the lack of contextualization and emotional connection of mainstream language learning apps.

LingoComics allows learners to curate their own stories based on situations and context settings, in which learners might be interested in role-playing. Beyond that, the app also provides activities for learner to test their contextual understanding, where they can piece together the story in a simple drag-and-drop activity connecting the narrative to their respective comic-style panels.

1.2 - Target Audience

The target audience for LingoComics is people who are learning English. Ideally, foreign language learners who are pre-intermediate level or higher. LingoComics is most helpful for learners who want a different approach from the more traditional curriculum-based method, through a relaxed and creative storytelling method. The story generation feature, called Story Designer in the app, was designed for learners who want to create a story that interests them in role-play, where a situation premise, setting, and main character can be described and integrated as part of the story's narrative and illustration.

2 - Background

Language learning apps traditionally focus on pre-defined curriculums, memorization and vocabulary-based learning. Mainstream apps such as Duolingo (Duolingo, n.d.) and Memrise (Memrise, n.d.) focus mostly on memorization and vocabulary building. These apps have their goals, and they help language learners learn a new language to varying degrees. While apps like Duolingo and Memrise serve their own purposes and work for some, LingoComics brings a new way to learn languages through storytelling. Having an immersive experience of going through a story that the learner can relate to can have a lasting impact on their ability to learn and pick up the new language.

Despite the success of these conventional methods, there has been a growing interest in exploring alternative methods that cater to different learning preferences. The new methodology acknowledges how individuals assimilate information and seeks to make learning more engaging and personally relevant.

2.1 - LingoLand: Immersive Learning

Several applications have begun to explore the intersection of AI and immersive technologies to enrich the learning experience. There have been a multitude of different approaches used for a variety of applications that provide the same methodology similar to LingoComics. These applications use artificial intelligence (AI) to leverage immersion and engage users in their world. For instance, LingoLand is an application that aims to *increase motivation, satisfaction, and learning success by mimicking real-life scenarios* (Seow, 2023). The use of AI in LingoLand serves multiple purposes – from personalizing the learning journey to dynamically adjusting difficulty levels of the language tasks based on the learners' performance. LingoComics shares many of its features with LingoLand, and both are pivotal in shaping an innovative method of learning languages. This section goes in depth about LingoLand's features and showcases its significance in using AI for language learning.

2.1.1 - Personalized Immersion

Although there are many similarities between LingoLand and its *competitors*, LingoLand stands out by its use of AI to engage learners by having an immersive game world. LingoLand aims to create a dynamic virtual environment designed to mimic real-life scenarios. Their approach addresses common barriers such as anxiety and lack of access to immersive settings by providing a relatable and controlled space where learners can practice their language skills at their own pace (Seow, 2023). LingoLand is an 8-bit game world of which your player character is a part. This world hosts NPCs (Non-Playable Characters) with their own personalities and lives. Your character's job is to interact with these NPCs and accept personalized missions, which amount to language-learning tasks. The player can accept numerous missions and go through them at their own pace. Then you receive instant feedback through voice-based interactions. This level of dynamic interaction and adjustment of difficulty ensures that learning remains challenging yet rewarding, which naturally sustains engagement and motivation.

2.1.2 - Generative World

The main feature of LingoLand is its use of generative AI capabilities. This technology is used for creating a very dynamic and customized learning environment through missions that aim to always keep learners engaged. Interacting with game characters, each with their own lives and personality in the world adapts to the learners' proficiency. LingoLand's AI technology is powered by Claude 2 developed by Anthropic, to generate its character personalities, contextually relevant missions, and responses. Every character has their own history, and subsequent interactions with them are based on historical information stored from generated details by Claude. With each character having their own personality, each encounter is unique, each mission is unique, and every journey; is unique, this level of detail to its world makes it a prime example of immersion, which helps a language learner be more engrossed in what they are learning, because each encounter has meaning.

2.1.3 - Encouraging Environment

LingoLand's AI virtual environment is important in creating a supportive learning environment. The characters in the game are designed to encourage and provide learners with

confidence in their current language skills and promote failures as a way for healthy learning. Promoting *healthy* failure for learners enables them to be confident in their learning ability all the while being immersed in the game.

2.1.4 - Conclusion

LingoLand demonstrates the significant advancement in language learning tools. By leveraging AI to create personalized missions and an immersive world, it addresses common challenges language learners face, such as motivation, memory retention, and anxiety. The app's focus on creating an encouraging environment while having a dynamic and immersive world helps set a new method for what AI technology can bring to language learning education.

2.2 - Storyfier - Learning Comprehension

Although immersion engages users in language learning, learning vocabulary is just as important in learning a language. Storyfier aims to offer a more tailored experience for learning language through target words in a small story, which deviates from the traditional methods seen in other mainstream apps. The core feature of Storyfier is generating stories that incorporate learners' target words (Peng et al., 2023). This approach is set to improve memorization and understanding of vocabulary through the medium of a story, which provides a natural setting for learning. Storyfier's learning process can be separated into three sections:

1. Getting users to read generated stories while covering target words
2. Conduct a cloze test
3. Use target words to write a new story with adaptive AI assistance

Each method engages learners passively and actively using new vocabulary (Peng et al., 2023). This three-stage learning process is designed for memorization and integration into daily vocabulary, effectively increasing their verbal vocabulary.

2.2.1 - Reading generated stories while covering target words

The first stage of Storyfier's learning process involves having users read generated stories while covering target words that will include the vocabulary they are trying to learn. This

process aims to provide a contextual learning environment, in the paper, they state that *vocabulary is learned best in a contextual learning environment* (Peng et al., 2023). The benefit of this approach is that it supplies learners with a connection to the material they are reading, thus creating a purposeful flow (Peng et al., 2023). Storyfier uses existing material for feeding their LLM to create relevant stories, for example, newspaper, articles, magazines, and novels (Peng et al., 2023). This method creates a rapport with their participants by having them read the relevant information, as they are reading a narrative that they will remember.

2.2.2 - Cloze test

The following reading phase requires learners to perform a cloze test, which is a test that involves filling in the blanks of a piece of text with an appropriate word. In Storyfier, this improves the learner's comprehension and retention of target words by requiring them to fill a sentence with an appropriate word in a similar context to the first phase where they initially encountered the word. In Storyfier the idea for the cloze test is simple, they will blank out words in the narrative text from the first phase with the target words, and it's the learner's objective to fill the blanks in with the appropriate word (Peng et al., 2023). Since the cloze test is a common vocabulary learning strategy employed in textbooks for example, the focus for the cloze test in Storyfier is on *target words* in a relevant context for *language-focused learning activities* (Peng et al., 2023).

2.2.3 - Use target words to write a story with adaptive AI assistance

The final step in the learning process challenges the learners to actively use the target words by writing a new story. Storyfier's adaptive AI assistance supports this phase, reflected in the UI. The UI has an integrated AI system that will check the grammar of the story being formed, lexical diversity and coherence (Peng et al., 2023). This innovative system focuses on enhancing the learner's understanding and memory of vocabulary through the active use of target words. The AI's constant feedback on the story's lexical diversity, coherence and grammar has the user learning at every step. The adaptive AI is geared towards the learner's current proficiency level, providing a personalized learning experience for each individual.

2.2.4 - Conclusion

Storyfier's learning process is designed to provide a new and innovative approach to vocabulary learning, moving away from traditional memorization techniques to leveraging the use of AI in language learning. By guiding learners through each of its phases of vocabulary learning, the use of AI technologies supports comprehension and retention of newly acquired vocabulary.

2.3 - Exploring the Benefits and Challenges of AI-Language Learning Tools

While it's important to analyze applications that have used AI to have an engaging language-learning experience, it is crucial to understand the benefits of these tools. The paper, *Exploring the Benefits and Challenges of AI-Language Learning Tools*, (Rebolledo Font de la Vall and Araya, 2023), examines how AI language learning tools can benefit language learners. For instance, the paper outlines the benefits of AI language learning tools, such as efficiency in learning speed and accessibility. However, the authors also note some challenges, such as the lack of human interaction and difficulties in *replicating cultural and contextual nuances* of a language (Rebolledo Font de la Vall and Araya, 2023).

2.3.1 - Benefits of AI-Language Learning Tools

The increasing use of AI for various tasks has proven beneficial for users, which is now evident by its recent advancements. Benefits can be in the form of efficiency, accessibility, and cost-effectiveness. This study highlights that efficiency is a benefit that language learning tools have, because it can provide immediate feedback to learners, which therefore helps them improve their skills faster (Rebolledo Font de la Vall and Araya, 2023). Since most people worldwide have a mobile device with internet access, they can also access these AI language learning tools, allowing learners to learn at their own pace from any location. Another benefit that these tools have is their cost-effectiveness, since many AI language learning tools are free or low cost, it makes them a more affordable option as opposed to traditional language learning methods. These benefits are important to consider when developing AI language learning tools.

2.3.2 - Challenges of AI-Language Learning Tools

Although AI language learning tools have benefits, some challenges are present. This paper outlines potential challenges of AI language learning tools, for example lack of human interaction and cultural and contextual related issues of languages. The most important limitation or challenge with AI language learning tools is the need for more human interaction (Roxana. R, Fabian. A., 2023). Since AI tools interact through a device, more human interaction is needed to make it easier for learners who prefer a more personalized learning environment. The cultural and contextual differences between languages are often overlooked along with limited human interaction. Since language learning is not just about learning the “core language,” but also about understanding how cultural and contextual factors contribute to language learning, AI tools have difficulty replicating such scenarios (Roxana. R, Fabian. A., 2023). Thus, while AI language learning tools offer an innovative approach to education, their effectiveness is hindered by these limitations.

2.3.3 - Conclusion

To conclude, AI language learning tools represent a leap forward in educational technology, offering advantages in terms of efficiency, accessibility, and cost-effectiveness. This paper overviews how these tools allow users to learn languages more efficiently than traditional methods. Though, AI tools have their benefits, some challenges arise, such as difficulty in contextual understanding and limited human interaction. Understanding the benefits and challenges of AI language learning tools, LingoComics aims to use the benefits to its advantage and address the limitations of having a different approach to AI-based language learning tools.

3 - LingoComics

LingoComics is a language learning application designed to engage users in the language learning process through context comprehension and interactive storytelling. Using advanced artificial intelligence technologies, LingoComics creates unique, contextually relevant narratives that mirror real-life scenarios, generating a natural connection between the user and their learning journey. This innovative approach ensures that each story engages and educates the user by embedding language learning with everyday situations.

LingoComics is a full-stack web application, this chapter describes the architecture, design, and implementation of its two main components: front-end and back-end. The back end includes prompt engineering and integration with OpenAI APIs, and it composes the majority of the technical contributions of this thesis. In the front end, LingoComics' user interface (UI) and experience (UX) build on the familiarity of existing systems such as Netflix to simplify the learner's interactions (Netflix, n.d.).

From a learner perspective, LingoComic has three main features:

1. **Story Designer** (Content curation): This module allows learners to design their own stories with creative freedom that they can relate to.
2. **Story Activities** (Practice): When interacting with existing stories, learners can choose between two activities: StoryConnect and Create-your-own-Adventure.
3. **Achievements** (Progress tracking): Displays the progress from StoryConnect, showing metrics such as points accumulated and time played for each story.

, From the homepage, learners can create or interact with a story. A story is composed of the following concepts:

Scenario: A scenario includes the initial conditions or description of the story.

Situation: A situation is the first part or chapter of the Scenario, which can have multiple situations. A situation is essentially an *arc* or *chapter* that happens within the story.

Moment: Moments are individual and distinct events that happen in a situation. Each situation has four moments.

Character: The character is one of the main focal points of the Scenario. This character can be edited by the user in their own personal stories.

Setting: A physical location or atmosphere of the Scenario. The setting can be edited by the user in their own personal stories.

Images: Within LingoComics, images are set to be comic-style, reflecting the application's name.

LingoComics is a framework for interactive storytelling, which is divided into 3 basic parts: the main scenario setting, the premise, a series of situations that happen within the scenario, and 4 key moments within each situation that move the narrative forward. This 3 phase structure was designed with a UI-first mindset, ensuring the language learning experience is intuitive and coherent. The culmination of this design is a drag-and-drop quiz activity, where the user's job is to match text to images based on contextual clues in the text, which reinforces language comprehension in an immersive fashion.

To add, LingoComics allows users to craft their own stories, which enables personalized learning experiences that relate to them on a personal level. This feature uses the 3 phase story separation to ensure the user is not overwhelmed by the interface. With the user at the center of a narrative, LingoComics attempts to bridge the gap between traditional and immersive language learning methods. By leveraging artificial intelligence with interactive storytelling for learning a language, LingoComics diversifies the landscape of the language learning sphere.

3.1 - System architecture

LingoComics design follows a learner-driven or user-driven approach. Ultimately, the key to creating an engaging application is the general user interface and experience. A minimal yet effective design for LingoComics was essential as it will determine whether the user is invested in the app. A minimalist approach would make LingoComics friendly for new users due to the ease of use which would result in their engagement with the application. This section reviews

all the tools and design philosophies used to create LingoComics. The general system architecture can be seen in Figure 1, which represents the backend and the API services used there and the data load and frontend and how they interact.

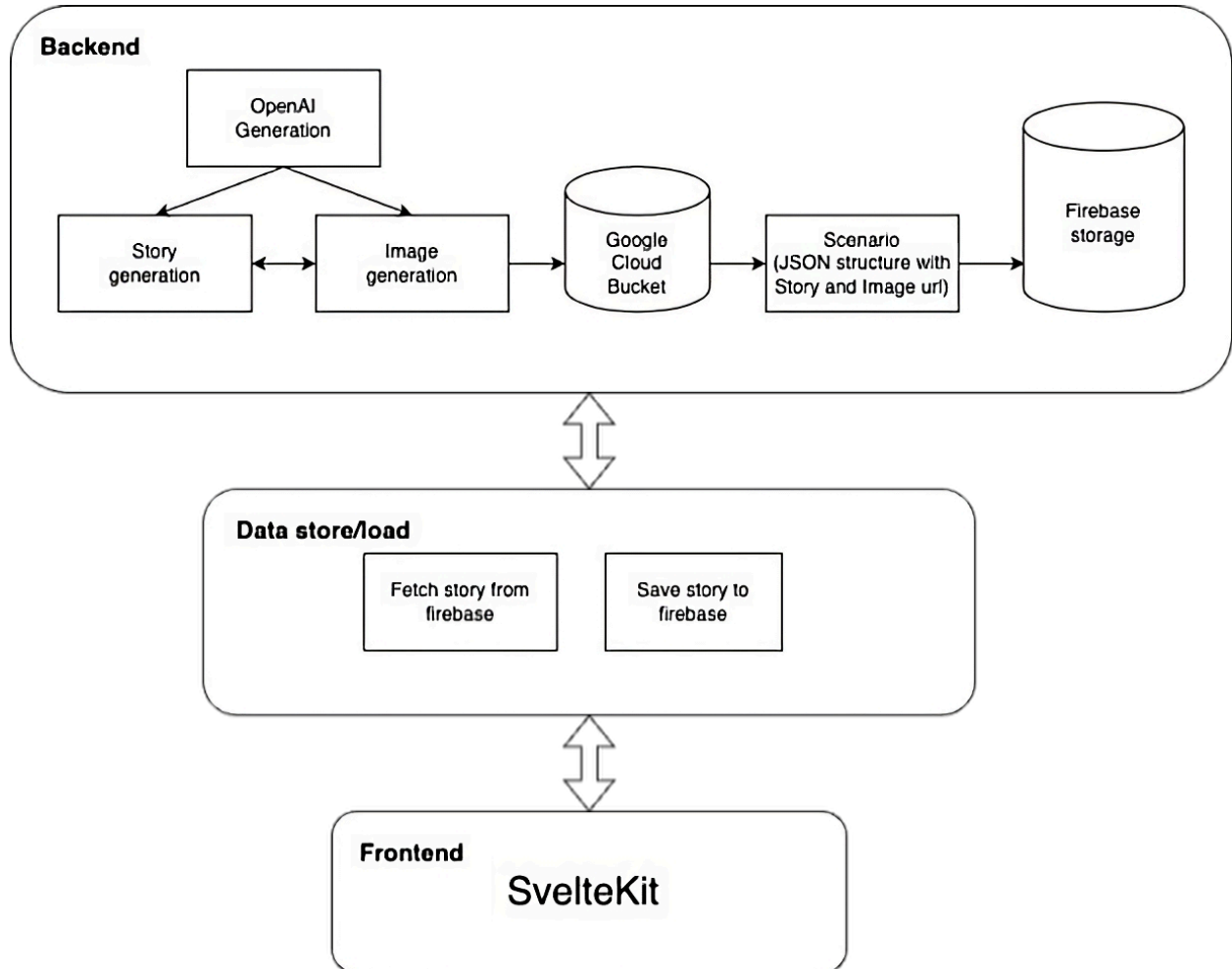


Figure 1: Application Architecture/Framework

3.1.1 - Wireframe

The initial step for creating an engaging and effective user interface for LingoComics involved the strategic wireframing of the app's prototype. The inspiration for LingoComics was the sleek and minimalist design of Netflix, and many of LingoComic's design choices are influenced by Netflix's design, as the ease of navigation flows well with the idea of LingoComics. Figma, was the tool of choice for bringing to life this minimalist design, as Figma's toolkit allows for an effective way to incorporate most design features required for LingoComics (Figma, n.d.).

The initial wireframe was to create a simple home page and stick to the theme of Netflix, for every scenario is shown as a long bookmark-type card, as seen in Figure 1. The tabular cards for each scenario are reminiscent of Netflix's design and are carried out throughout the app. The main focus for creating this design was with an ease-of-use design philosophy in mind, at a glance the user should be able to understand how this app functions.

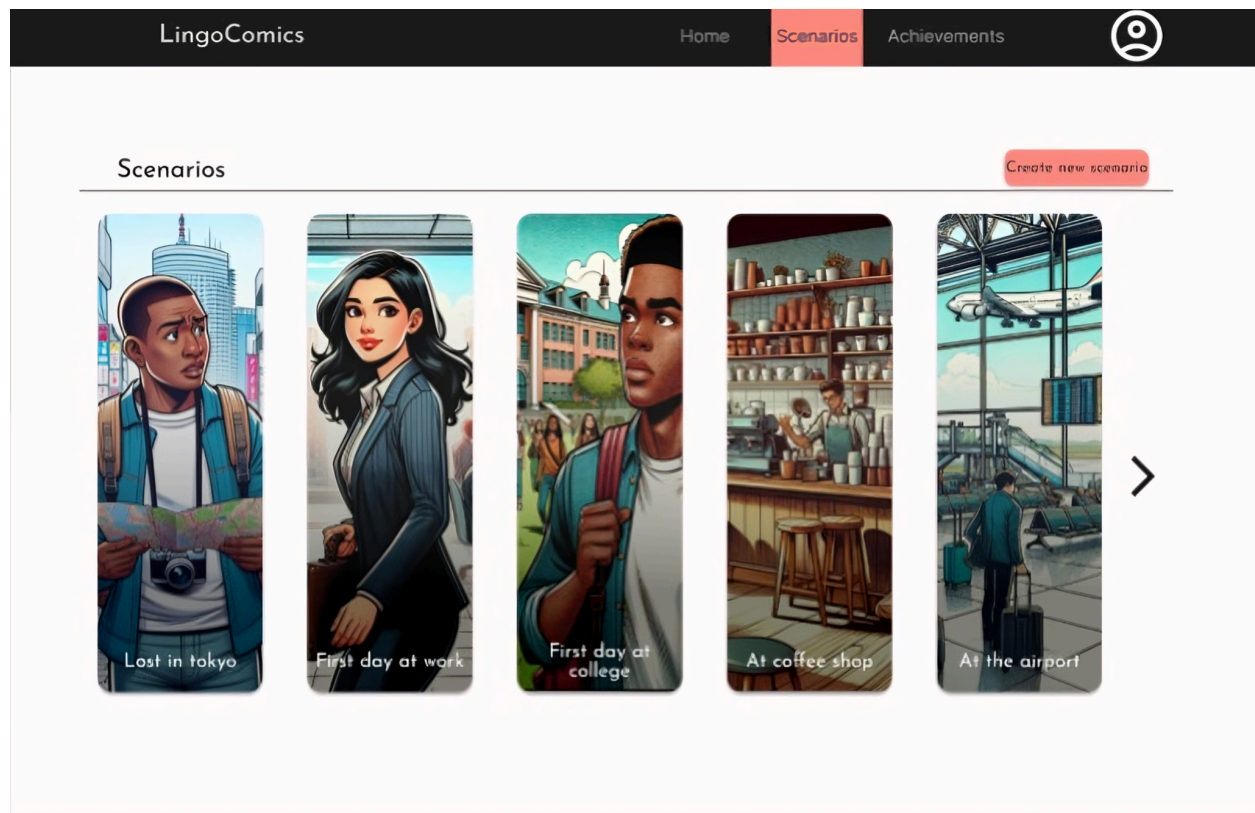


Figure 2: LingoComics homepage wireframe

The next step was to create a quiz layout for the app, as that is where LingoComics user interface/user experience (UI/UX) is vulnerable, as there are many moving parts involved in creating a drag-and-drop interface with text and multiple images. The idea for this design was sticking to the comic theme of the app, and having the layout be that of a comic book. A comic book-style design effectively created an engaging interface for a drag-and-drop layout, as seen in Figure 2. This screen includes showing the scenario's title and situation at the top, and the screen is split into 2, where the left side will show the images and the options are shown on the right side. Above the images are pagination dots that fit the theme of a minimalist design.

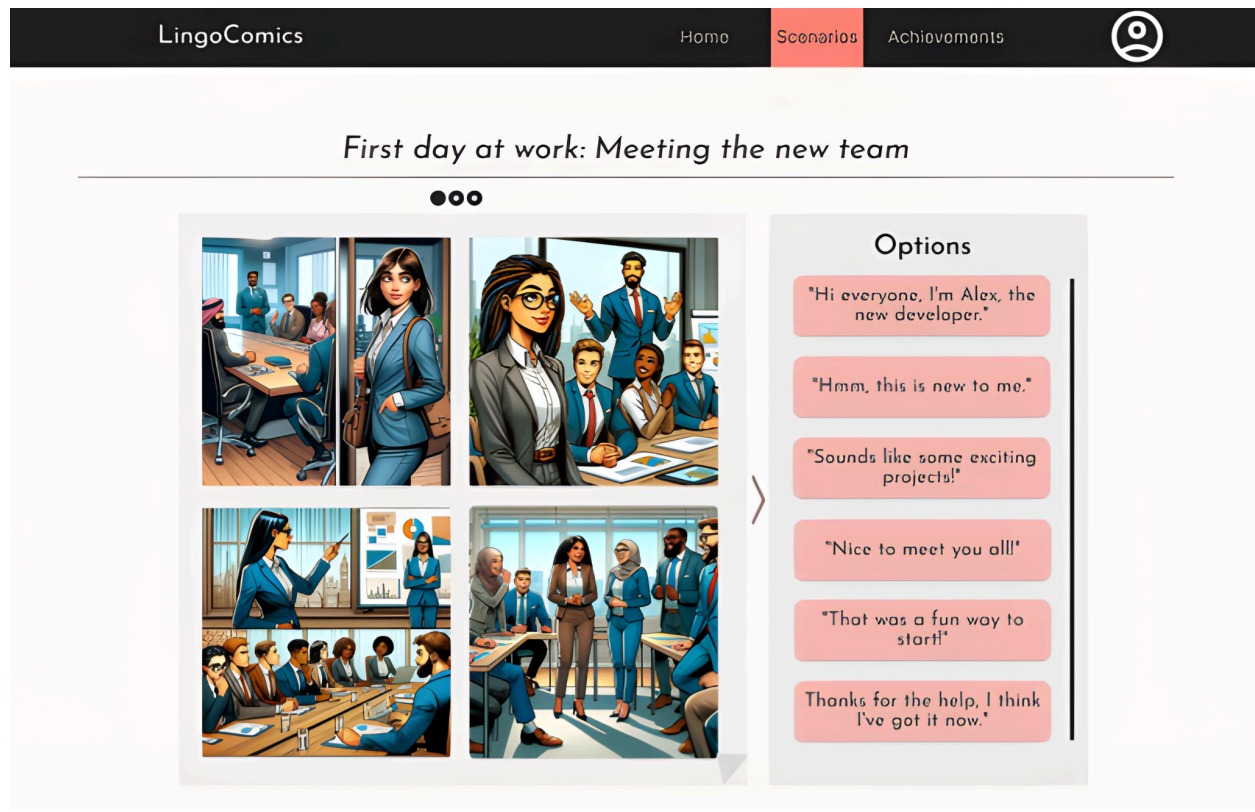


Figure 3: StoryConnect (drag-and-drop mode) wireframe

3.1.2 - Tools

After generating the wireframes, it was crucial to pick tools and technologies that would aid in developing LingoComics. This section covers all the tools, technologies, and APIs (application programming interface) used to create LingoComics, and how these tools are woven together to create the best language learning experience.

3.1.2.1 - Sveltekit

SvelteKit (SvelteKit, n.d.) is a UI framework built on top of Svelte, which uses a compiler that enables the creation of relatively simple to complex applications with minimal work compared to other front-end frameworks. The efficiency of Svelte in its compiler allows for the rapid creation of apps. SvelteKit also has support for TypeScript, meaning that most code errors would be caught before deploying or viewing the webpage and supported the rapid development of LingoComics. With LingoComics, using Svelte meant creating very minimal code that reloads very fast, and this was useful as LingoComics uses a lot of images and some

animations, which meant that using SvelteKit would natively handle many issues around using images and animations in apps when working with other frameworks.

3.1.2.2 - Firebase

Firebase (Firebase, n.d.) is a development platform developed by Google (Google, n.d.). For LingoComics, Firebase's Firestore service was used, as Firestore allowed for rapid prototyping and storage of data. Since Firestore complements the standard web JSON format, storage and retrieval of data was very quick.

3.1.2.3 - Google Cloud Service (GCS)

Google Cloud service is a cloud service that was developed by Google (Google, n.d.). This service hosts a plethora of microservices and APIs, such as text-to-speech APIs, image recognition, speech recognition, and more. The main service for LingoComics was using a cloud-hosted storage bucket for storing the images, as this was the fastest way to fetch images for the image, minimizing load times considerably. The bucket for LingoComics stores a public link for each image which is then directly referenced when retrieving the image in the fetch process.

3.1.2.4 - OpenAI API

Open AI API (OpenAI, n.d.) is an API service developed by OpenAI (OpenAI, n.d.). With the recent advancement in artificial intelligence technologies, OpenAI provides many interesting AI services, such as text generation, image generation, image vision, and many more. In LingoComics, the use of text generation and image generation was crucial as it's the main point of generation for all of its content. The text generation was responsible for creating the overall story for each scenario, and this service was also leveraged to create descriptions for images, which were then directly fed into the image generation API.

3.2 - Possible use case

Imagine yourself going on a vacation to Tokyo, a city and culture that you are unfamiliar with. You are excited to visit Tokyo but feel anxious -- thinking about how you will navigate through the metropolis with your limited knowledge of their spoken language. This is where

LingoComics comes in, as it can bridge the gap between your anxiety and the excitement of visiting a new country.

You think of possible situations that may happen to you and realize what would happen if your phone ran out of battery and you were left alone to navigate where you were trying to get yourself. You might think that this is a scary situation, as you do not know the language spoken in Tokyo, and the metropolis of Tokyo may be hard to navigate. What would you do if this were to happen to you? These thoughts further feed your anxiety.

This is where LingoComics comes into play with its *Story Designer* feature, which can help turn your fear and anxiety into anticipation. Through this tool, you can craft this very scenario, and place yourself in the situation that you fear. You can describe and alter the situation to fine details, such as your appearance, the setting, and each moment that you may encounter. *The Story Designer* will then generate this narrative with comic-style images that mirror the situation that you described. Using the Story Designer, you can simulate an experience of overcoming the anxiety of navigating Tokyo and tackling the language barrier, all within a safe virtual environment.

Through this personalized experience, LingoComics can teach you more about language, but it even prepares you for the real world by creating contextually relatable scenarios. This is not only limited to visiting another country to overcome a language barrier but it can be applied to relieve anxiety for some tasks that may invite such anxiety, for instance, ordering at a restaurant. With LingoComics, you can confidently explore encounters as it can help you learn a different language with a relatable situation which will in turn help with retention.

3.3 - Story designer

Creating the Story Designer was a key component in LingoComics as it aims to create an innovative story-designing feature. This feature has two main primary functionalities:

1. Contextually relatable story generation (OpenAI's GPT-4-turbo text-completion)
2. Comic style image generation (OpenAI's DALL-E-3 image generation)

The way users can engage with these features is by interacting with a simple form that will prompt the user to go over the premise of the story, setting, number of key situations, emotional tone, and primary challenge or conflict. The premise of the story can be as simple as stating, “At a restaurant,” while the setting is a general location, “Fast food restaurant”, to a more specific location such as, “The City of Toronto,” then the app defaults to having one situation that will have with it four key moments, and the user can either specify or leave the emotional tone and primary challenge blank. After the user is satisfied with their specifications, they can click the “Generate” button, which will begin the story designer’s main pipeline.

3.3.1 - Overview of contextual story generation

After the user has entered all their desired inputs into the form, the pipeline for story generation is started. The first step in the pipeline for the story designer is crafting the contextually relatable story powered by OpenAI’s **gpt-4-turbo** text-completion API. The pipeline begins by generating an overall story, including all the user’s specifications entered in the form. The prompt that crafts this story is as follows,

“Generate a story that is contextually relatable to everyday real-world experiences. The title should reflect common situations, such as ‘First day at work’, ‘Lost in a city’, ‘First day at school’. Use the following inputs to inform the title, ensuring it aligns with these themes:

- Premise: \$
- Setting: \$
- Tone: \$
- Conflict

The story should evoke a scenario that is easy to visualize and relate to. Such titles include ‘Eating at a restaurant,’ ‘Meeting new neighbours,’ or ‘Exploring a new city.’ Avoid overly specific or niche scenarios to maintain broad relatability. The story should be relatable, for example, a person in a story struggling to order a coffee, or talking to a stranger, or being lost in a city”.

This prompting style focuses on the few-shot prompt engineering technique, enabling context learning by giving existing logic within a specific domain. For instance, in our example, the reference to ‘First day at work,’ ‘Lost in a city,’ and ‘First day at school’ were used to generalize the context of the overall story, so that it remains broad and relatable to these premises

(Prompt Engineering Guide, 2023). This prompting style helped generate stories that remain relatable and not too specific or outlandish. The next step was to summarize the story, this was

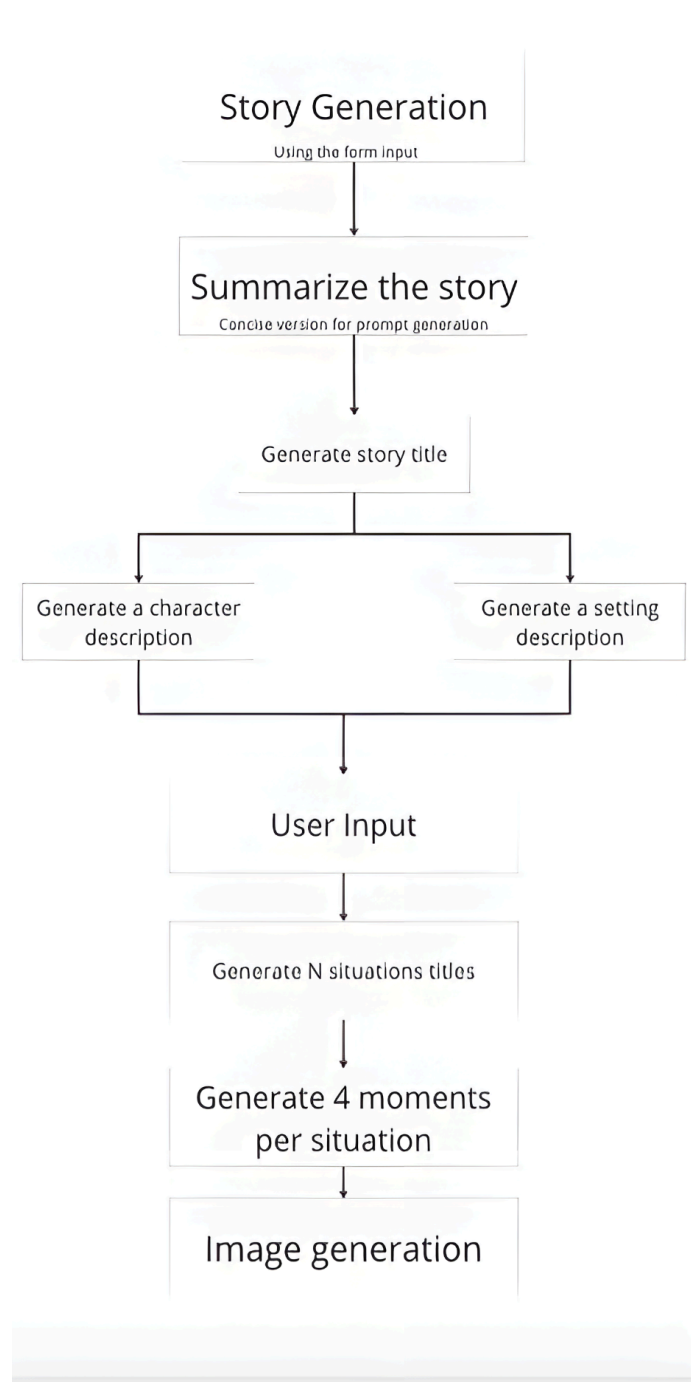


Figure 4: Story Generation pipeline

done to later feed into the image generation sub-pipeline. The next step will extract the story title from the overall story generated.

Now the character description is extracted, and immediately a new and more detailed character description is generated based on who the story decided to generate, basically expanding on the physical appearance of the character which plays a key role in the story generation pipeline as the appearance helps the image generation create a consistent character without deviating too much from the original image. The next stage in the pipeline generates the situation title(s), and the four key moments that happen as part of the situation.

After everything is generated, the story is displayed on the screen in a timeline representing the generation. First, the character description is shown where the user can edit the description, editing the description will fire up the pipeline mid-way through which will update the character in the overall story. The next item will present the setting which gives a detailed description of the setting, which can also be edited, and it also follows a similar re-generation to make the story adhere to the new and/or updated context of the story. The next panel will display the situations and their respective key moments which are also editable; editing the moments results in yet another re-generation that engages in an alternate *shadow* pipeline that updates the story's context from the referenced point onwards. If the user is satisfied with the content that was generated, they can now engage with the next functionality of the story designer which is the image generation process.

3.3.2 - Comic style image generation

With all textual content generated, users can now generate images for their stories. All images are generated using OpenAI's DALL-E-3 API, which makes it possible to create high-definition and expressive images. The image generation step begins when the user clicks on the "Generate images" button in the UI, where the sub-pipeline for the image generation starts. First, the scenario image description is generated, this is done by providing the summary of the scenario, the character description, and the setting to a preliminary function. The prompt for this text generation focuses on detailed descriptions:

"Generate a concise summarization of the following scenario: **\$scenario**.

The description of the setting is: **\$setting**.

Create a single comic-style cover image without typography that will capture the essence of the overall scenario, without any text or speech bubble. The artwork should capture the essence of the scenario WITHOUT including any text or words in the image.

The scenario should be the focal point of the image with an environment that has the character: **\$characterDescription** in it. Ensure the image is dynamic and conveys a sense of narrative or action related to the scenario. THE IMAGE HAS TO BE COMIC-STYLE."

With the description of the image being made, the image is then generated using OpenAI's DALL-E-3 image generation API. The scenario summarization was provided as the scenario parameter in the function due to the hard limit of DALL-E-3 with its 1000 token limitation, this required pre-generating concise summarization for the scenario. Over the 1000 token limit, the image fails to generate, causing the current image not to be generated. After the scenario title image is generated, the situation images are generated, this is done in a simple for loop, following a similar process to the scenario image generation. Then for each situation, four-moment images are generated which also follow the same process as the previous images, this is done in an inner for-loop within the situation for-loop. Figure 5 below shows the general structure that the image generation follows.

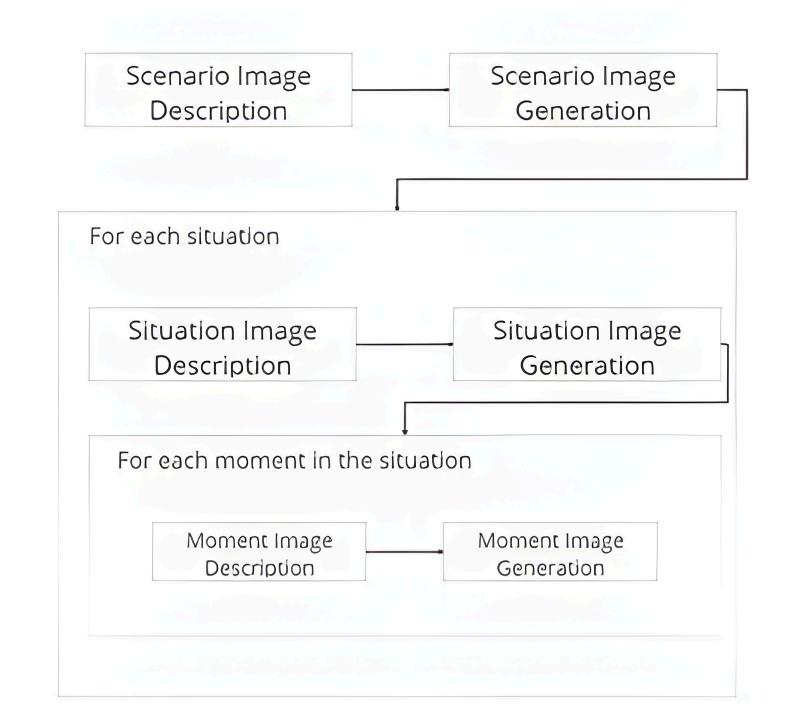


Figure 5: Image generation pipeline

Generating images is a time-consuming process and generating a minimum of six images for a scenario with a single situation can take between 3-5 minutes. Ideally a stream process would be used to hydrate the frontend as soon as an image is ready, keeping the user engaged on the app. After all the images are generated, the UI is hydrated with all the images. The user is presented with the main scenario image that has a larger card on the left and the situation images are on the right in a banner-like card (Figure 6).

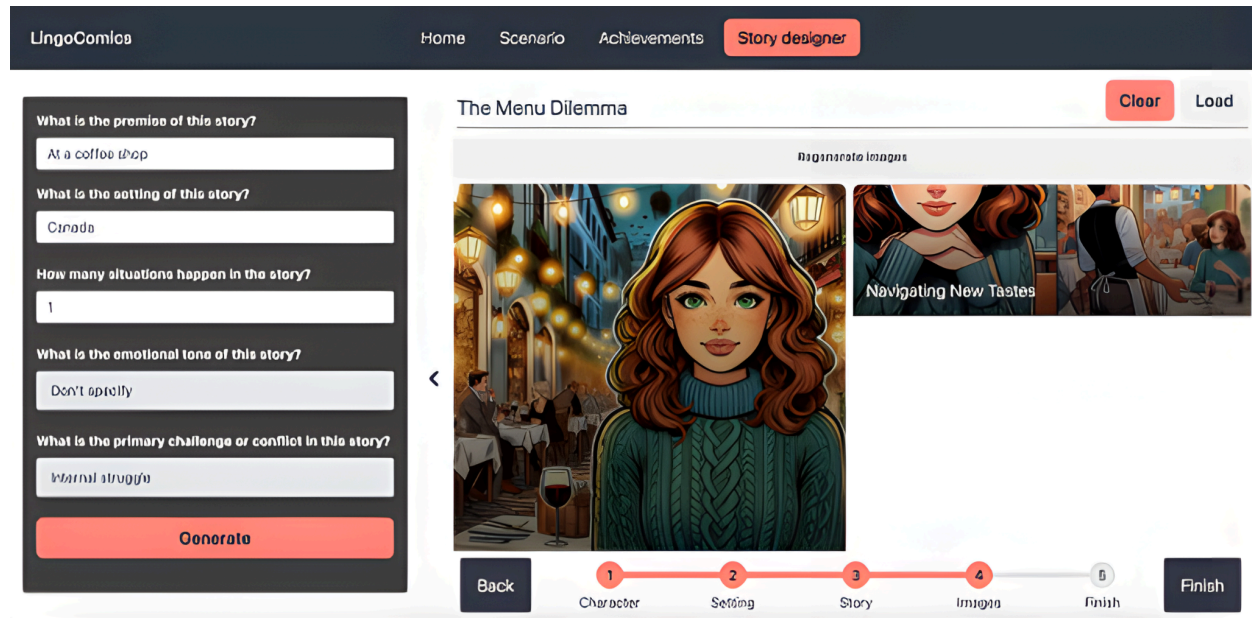


Figure 6: Story Designer; Step-4 Image generation screen

The situation image(s) are interactable. Clicking or tapping on the situation image makes it the main image, taking the spot of the scenario image. All the moments for the situation are presented on the right side in a 2x2 grid (Figure 7).

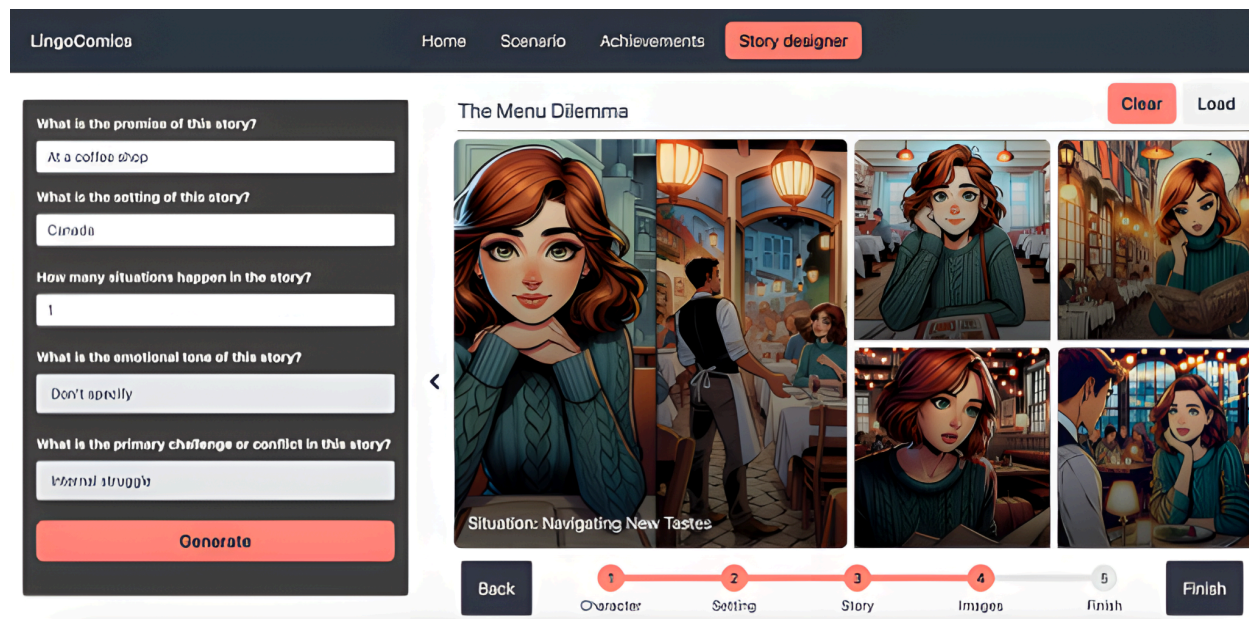


Figure 7: Story Designer; Step-4, Situation Images

If the user is not satisfied with the image, a “Regenerate images” button is provided, which will give a warning message stating that this action will overwrite the images you had previously, and also state the cost of generating these images.

3.3.3 - Challenges of creating a coherent story

Developing a robust pipeline to generate on-demand dynamic stories was challenging. Generating consistent and coherent stories and images was a priority for LingoComics, as without it, the core of the app would fall apart. Along with consistent and coherent stories, the limitations of the API also proved challenging, as they initially hindered the quality of the output.

3.3.3.1 - Consistency in image generation

Generating a consistent story is a fundamental aspect of storytelling, keeping the user engaged with the content. With LingoComics, the approach was different than most forms of storytelling, as each story is separated into three parts, the scenario, which is the premise, the situations,

which are the subplots, and the four key moments for each situation which are distinct action or event that take place within its respective situation. With the text generation pipeline, the overall story structure generated had a coherent plot throughout every three parts. The issue was introduced when generating images, as each image reflected the individual parts, scenario, situation, and moments, and the consistency between these images was lackluster, due to each image being generated separately.

Separately generating images required pre-generating physical descriptions for the character and setting. Since images are generated separately, the seed for each image is different, so achieving true consistency was extremely difficult. For each image, the character and setting would change, sometimes even changing the style of the image (Figure 8). This introduces complexity when going through the story as the character would change and most often have the same expression for each image, even when explicitly stating the emotion and situation that the character should be in. To overcome this, the image description was introduced to the text-generation pipeline, which described the character's physical appearance and setting. With the introduction of this extra layer, the images become consistent with about 80% accuracy in character and setting parameters (Figure 9).



Improved character and setting consistency

3.3.3.2 - API limitations

Another issue that challenged the development of LingoComics was OpenAI's token limitations, specifically the image generation with DALL-E-3. Currently DALL-E-3 only accepts about 1000 tokens for generating images. At each step of generating images, the prompt size grew due to the simple cascading technique to build on the story, this proved to be a challenge with LingoComics. With each image further in the story, more context from the previous step of the story was needed; when generating the moments the prompt grew too large to be able to feed

into DALL-E-3. To overcome this obstacle, the text generation pipeline was introduced with the summarization of the scenario, and a concise description of the character and setting reduced the token count by ~75. This was enough for 99% of images to be generated.

3.4 - Interactive activities

The main objective of LingoComics is to create an engaging, interactive language learning experience. This meant giving the users the creative ability to craft their own stories by having a language-learning mechanism. LingoComics hosts two interactive activities: the drag-and-drop quiz and the *create-your-own-adventure* activity. The drag-and-drop activity is an exercise that requires the user to match contextual sentences with their corresponding images. It requires learners to understand the context that is being presented in order to accomplish the task. On the other hand, the *create your own adventure* activity allows learners to take control of their language learning journey. Enabling learners to construct their own stories, creates a deeper connection with the material, reinforcing memory retention. This section will take an in-depth look at both activities.

3.4.1 - StoryConnect

The drag-and-drop activity is a simple matching game that will test the user's ability to draw a connection between text and image. As part of LingoComics' structure, each story is separated into 3 distinct parts: the scenario, the situations, and four key moments for each situation. With that in mind, the user is presented with four key moments for each situation starting from the first situation of the story. The user's job is to read the options on the right side, then drag and drop them to the image on the left they think best suits the description (figure 10). Each text option has highlighted keywords that can help the user understand more about the sentence and image, the keywords denote key nouns and verbs in the sentence. On the right side of each option is an info button, clicking on this button will open up a pop-up modal that will grant a descriptive look into the keywords. Each keyword has an explanation, and if applicable, the explanation will be contextually related to the sentence itself.

There is also a speaker button, that will allow the user to trigger the text-to-speech functionality. The text-to-speech functionality is enabled through OpenAI's speech API, which

generates a natural-sounding text-to-speech sound compared to other alternatives. After the user is sure of their guess, they can drag-and-drop the text to the image they think best fits the description, and they can click the “Check answers” button to check if the text you dragged onto the image is correct. The “Check answers” feature can be used one at a time or after every option is dropped to their images. If the answer is correct, a green border around the image will appear, otherwise, if the answer is incorrect, a red border will appear, and reset the position of the text option. This quiz mode also features a points system, for which for every image you get the correct answer, you are rewarded 25 points for each text-image pair, and an incorrect answer will be punished with 10 points for each text-image pair.

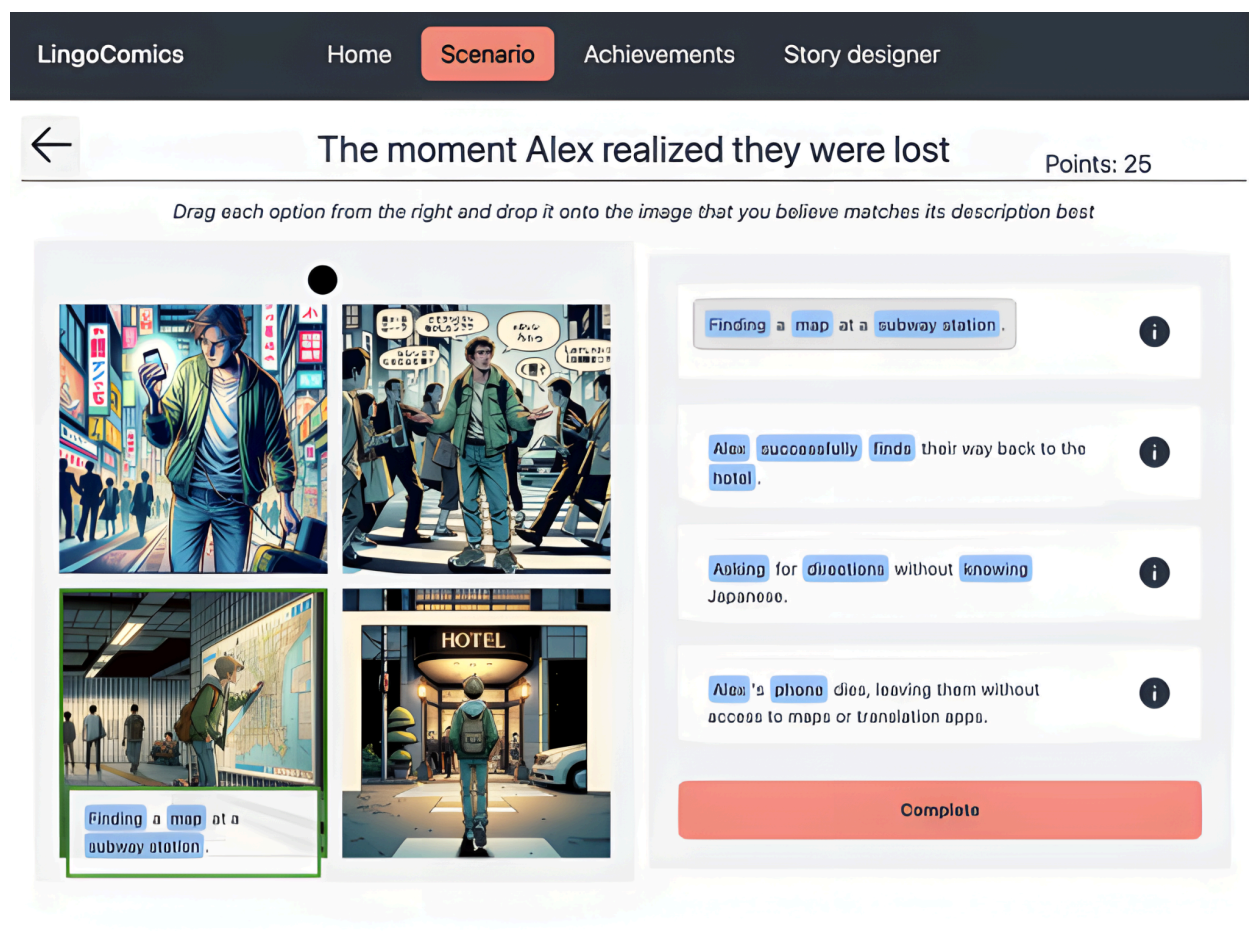


Figure 10: StoryConnect page

3.4.2 - Create your own adventure

“Create your own adventure” mode is another activity for learners to participate in, allowing them to craft their own stories. This mode exists for every scenario that is part of the LingoComics, for every scenario you can *create your own adventure*, which is based on the first situation for every scenario, and the learner can build off this. When starting this activity, you are presented with a similar layout to other features of the app; on the left, there is an image with a small narration to go along with what is happening, and on the right will be three options or branches you can select to move the narrative forward. The options on the right have a layout similar to the drag-and-drop quiz, with highlighted keywords and an info button that allows you to see each keyword in-depth, along with a text-to-speech button. Choosing an option will trigger a pipeline that generates the next part of the story based on the option selected previously. The pipeline is simple, each time an option is selected, the option, along with the narrative and the details of the scenario such as character description are carried over to generate the next set of options and images. This mode lasts indefinitely unless you interact with the “End adventure” button which will save your progress, which can later be viewed in the achievements tab. Like every other feature in LingoComics, this mode has an autosave system that will save your adventures in the Firestore database.

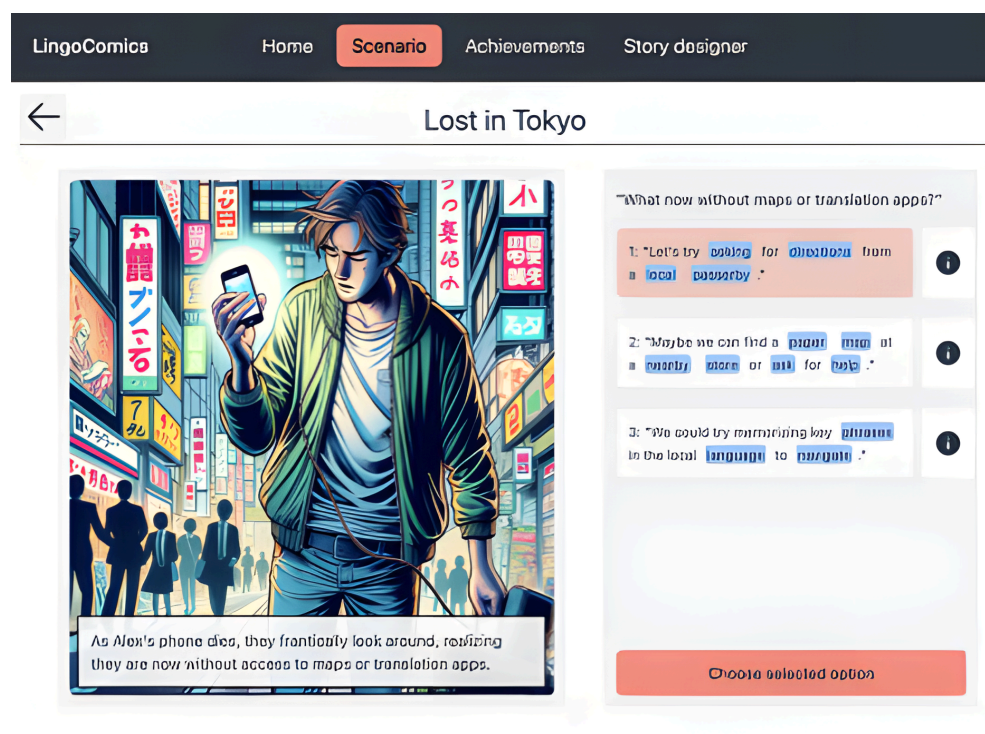


Figure 11: Choose-your-own-Adventure mode

4 - Evaluation

To verify the feasibility of LingoComics and its application to language learning an evaluation should be organized. An evaluation of LingoComics would assist in measuring its effectiveness as a language-learning application. This section will outline a basic evaluation plan that can be conducted for LingoComics, highlighting user evaluation through qualitative and quantitative measures.

For LingoComics' evaluation to be successful, it is essential to establish a set of guiding questions beforehand. These questions should then serve as a framework to assess whether LingoComics is useful as a language learning tool and is successful in user engagement. Some questions that should be asked are:

1. How does LingoComics help your language skills?
2. How easily are you able to relate to the stories contextually in LingoComics?
3. Are you able to retain new vocabulary and context clues after using LingoComics?
4. Does the user interface support the contextual storytelling learning environment?

With these questions in mind, the evaluation plan can be carried forward. To find the answers to these questions, there would need to be two sets of participants: one group of participants would use only the pre-built scenarios that are part of LingoComics, which can be group A. The other group would use only scenarios they make themselves using the story designer function, which can be group B. With the two separate groups of participants, each would use the app in a 1-2 week time period with periodic reminders for the participants to use the app. Group A would use only the pre-built scenarios to play the different activities: drag-and-drop and create your own adventure.

Group B would participate in the same activities, but instead they will use their own scenarios from the story designer. Since the points are recorded when doing the quiz we can use that as a quantitative measurement to see how the participants are doing. A logging functionality as part of the backend would also be beneficial to see if they are using features

such as text-to-speech. This basic evaluation plan would help answer the questions for both groups of participants, and drawing parallels between them can help with further development and improvement of LingoComics.

5 - Future work

LingoComics, in its current form, is a prototype for an innovative language-learning application that uses contextual storytelling for immersive language acquisition. To transform LingoComics into a personalized language learning experience, implementing a user account system for personal account creation is the initial essential functionality. With this system in place, users can view their own stories, which can be organized on the home page, they can also view stories created by other users, creating a sense of community engagement around the app. Furthermore, a user account enables future improvements to LingoComics, such as a comprehensive achievements system and a translation feature.

5.1 - Achievements

Building on top of the user account system, the achievements feature would be designed to generate further user engagement. To begin, the achievements system would have a trophy system that unlocks achievements based on how the user uses the app. An example of this is seen on the Quizlet website (Quizlet, n.d.), where having a user account would provide users with trophies/achievements for interacting with certain features within the app (Quizlet, n.d.). This would translate well into LingoComics by rewarding users with interacting with different app features. For example, achievements could be awarded for simply doing the drag-and-drop quiz, creating your own adventure activity, and using the story designer for the first time. This would be the first step in engaging the users with the app.

Another set of trophies can be attached to the number of points they have in drag-and-drop, for example, getting their first 100 points without mistakes and similar achievements. There would be milestone-based achievements as well, for instance, trophies would be given for the total accumulation of points at different milestones, such as a trophy for 100, 500, and 1000, for example.

This same methodology can also be used for the number of stories the user has created using the story designer mode. An achievements system for LingoComics can further help with user engagement and immersion, making their language learning journey enjoyable.

5.2 - Translations

A translation functionality would benefit LingoComics as it supports the language learning aspect of the app. A global language setting allowing the user to select a language that they are proficient in and the language that they are targeting to learn would be a way to enable translation within LingoComics. While saving both of these preferences, it can become clear which parts of the app should be in the language they are comfortable with and which parts should be in the target language. An example is the drag-and-drop and create-your-own-adventure activities with the highlighted keywords and text-to-speech features. Ideally, the highlighted keywords would be in their target language and otherwise for non-highlighted keywords.

This way the learner really has to understand the context of the text and be able to connect it to the image. The info button would enable the user to use the text definition of the keyword in their comfort language to understand what the keyword is, also, using the text-to-speech function would help with pronouncing the word. This level of interactivity would help users engage with the app on a personal level, by taking time to understand the context of each sentence and their connection to the image.

Another addition is to include a feature that reads the story out in a narrative format to the user after they have successfully completed the quiz. A translation feature at the core of LingoComics allows users to fully use its language learning features in a different format than other tools available.

6 - Conclusion

LingoComics is a language learning tool using AI at its core to generate contextually relatable stories that prioritize engagement and contextual understanding. With the recent advancements in artificial intelligence, using APIs that provide such services enables LingoComics to create these stories. LingoComics features a quiz mode that is based on a drag-and-drop matching framework, that has the user matching text options with images based on the context of the option. Along with the quiz, a story designer mode allows users to craft their own stories, and go through them at their own pace. Users can also take their own stories through different outcomes with the create-your-own-adventure mode, allowing them to branch to different narratives. The emphasis on a minimalistic user interface provides LingoComics with the ability to have many features that are easily accessible, thus reducing cognitive load and allowing users to focus more on the learning experience itself. This approach ensures that future work related to the app can be easily integrated, without overwhelming the user but effortlessly integrating it into the user's experience. Using a minimalistic toolkit such as SvelteKit and Firestore allows the fluid development of LingoComics. Additional elements can be added to LingoComics, for example, a detailed achievements system and a translation functionality. After having all these features in place, an evaluation can be held to determine whether this method of learning languages is engaging and feasible.

References

Duolingo. (n.d.). Mission. Duolingo. Retrieved April 12, 2024, from

<https://www.duolingo.com/info>

Figma. (n.d.). Figma. Retrieved April 12, 2024, from

<https://www.figma.com/>

Memrise. (n.d.). Learn a language. Memrise is authentic, useful & personalised.

<https://www.memrise.com/>

Netflix. (n.d.). Netflix. Retrieved April 12, 2024, from

<https://www.netflix.com/ca/>

Olivia, Seow. 2023. LingoLand: An AI-Assisted Immersive Game for Language Learning. *In Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct)*. <https://doi.org/10.1145/3586182.3625117>

Prompt engineering guide. (n.d.). Prompt Engineering Guide | Prompt Engineering Guide.

<https://www.promptingguide.ai/techniques/fewshot>

Quizlet. (n.d.). Captcha challenge. Retrieved April 12, 2023, from

<https://quizlet.com/latest>

Roxana Rebolledo Font de la Vall, Fabián González Araya. 2023. Exploring the Benefits and Challenges of AI-Language Learning Tools. *International Journal of Social Sciences and Humanities Invention*.

<https://valleyinternational.net/index.php/theijsshi/article/view/3771>

Zhenhui Peng, Xingbo Wang, Qiushi Han, Junkai Zhu, Xiaojuan Ma, and Huamin Qu. 2023. Storyfier: Exploring Vocabulary Learning Support with Text Generation Models. *In Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct)*. <https://doi.org/10.1145/3586183.3606786>