



## **CS3216 Assignment 3**

Artificial Intelligence Application

### **Team Members**

<b>Daniel Kok</b>	A0234683Y
<b>Tan Wee Kian, Justin</b>	A0252153N
<b>Lam Jiu Fong</b>	A0255851Y
<b>Ho Cheng En Bryan</b>	A0234509E

# Table of Contents

<b>Table of Contents</b>	<b>2</b>
<b>Milestone 0</b>	<b>5</b>
<b>Milestone 1</b>	<b>6</b>
Features Comparison	9
<b>Milestone 2</b>	<b>10</b>
Objectives	10
User Stories	10
Additional Notes	10
<b>Milestone 3</b>	<b>11</b>
1. Dynamic Content to Maximise Engagement	11
2. Personalised Feedback System	12
3. First Mover Advantage	12
<b>Milestone 4</b>	<b>13</b>
<b>Milestone 5</b>	<b>15</b>
Current Features	15
Future Features	16
<b>Milestone 6</b>	<b>18</b>
Pricing Strategy and Features of each Tier	18
Rationale For Freemium Model and Subscription Model	18
Revenue Streams	19
Factors Influencing Pricing Decisions	19
<b>Milestone 7</b>	<b>21</b>
Contextual Understanding	21
Adaptability	21
Multilingual Capability	21
<b>Milestone 8</b>	<b>22</b>
Example 1: Prompt for AI Response to Users	22
Example 2: Prompt for Initial Message for Context	23
Example 3: Prompt for Suggested Responses	24
<b>Milestone 9</b>	<b>25</b>
Choice of LLM and Provider	25
Considerations	25
Popularity	25
Cost	26
Language Support	27
Multimodality	27
Performance of Model	27
Conclusion	28
Model Setting	29
Max Tokens	29
Temperature	29

Frequency/Presence Penalties.....	29
<b>Milestone 10.....</b>	<b>31</b>
Product Name.....	31
Logo.....	31
Tagline.....	32
<b>Milestone 11.....</b>	<b>33</b>
Frontend/UI.....	33
Backend.....	33
Database.....	34
Testing.....	35
Hosting.....	35
Authentication.....	35
<b>Milestone 12.....</b>	<b>36</b>
Workflow 1: Interactive Conversation with AI assistant.....	36
Workflow 1.1: Speech to Text (User Inputs):.....	36
Workflow 1.2: LLM generates response to user inputs and text-to speech.....	36
Alternative: Text-Only Interaction.....	36
Conclusion and Rationale.....	36
Workflow 2: Selecting Scenario to Practice.....	36
Workflow 2.1 (Alternative 1): Fixed Scenarios only.....	36
Workflow 2.2 (Alternative 2): Customised Scenarios only.....	37
Conclusion and Rationale.....	37
Workflow 3: Receiving Feedback on Performance.....	37
Workflow 3.1: End-of-Session Feedback.....	37
Alternative 3.1: Continuous Feedback Updates.....	37
Workflow 3.2: Hints for Enhancing User Engagement.....	38
Alternative 3.2: Singular Suggestion (Previous implementation).....	38
Conclusion and Rationale.....	38
<b>Milestone 13.....</b>	<b>39</b>
User Control through Context Input.....	39
Providing Multiple Suggestions.....	40
<b>Milestone 14.....</b>	<b>41</b>
Landing Page Considerations.....	41
Implementing OGP.....	42
<b>Milestone 15.....</b>	<b>43</b>
<b>Milestone 16.....</b>	<b>44</b>
Product Name.....	44
Launch Date.....	44
URL.....	44
Product Hunt Tagline.....	44
Topics.....	44
Video.....	44
Description.....	44
First Comment.....	44
Images Display.....	45
<b>Milestone 18.....</b>	<b>47</b>

**Areas of Improvement..... 48**

**Annex..... 50**

    Milestone 9 LLM Comparison.....50

        Test 1.....50

        Test 2.....52

**Github Repository.....54**

**Hosted Application..... 54**



## Milestone 0


Our application is a **language learning platform** that focuses on immersing users in **real-world scenarios**. Each scenario is designed to provide **contextual learning**, where the flow of conversation is **dynamic**, driven by interaction between the user and a generative AI. This allows for **personalised** and **adaptive language practice**, offering a more **natural and engaging learning experience**.

Our team has identified the following problems which SpeakEasy aims to solve:

1. **Lack of Contextual Learning:** Traditional language learning methods like Duolingo and Babbel often teach vocabulary and grammar **in isolation or short sentences**. This makes it **difficult** for learners to **apply their knowledge** in real-world situations. SpeakEasy addresses this by **immersing users in realistic scenarios**, providing **crucial context for language use**.
2. **Limited Personalisation:** Many language learning platforms offer a **one-size-fits-all approach** and **cannot adapt to a learner's specific needs** or interests. SpeakEasy uses generative AI that allows for **dynamic, personalised conversations** that adapt to each user's skill level, interests, and learning pace.
3. **Boring Repetition:** Traditional language learning methods can become **repetitive and boring** with a **fixed set of lessons** and materials. By offering **dynamic, AI-driven scenarios and content**, SpeakEasy **keeps users engaged and motivated** to continue learning.

## Milestone 1

Competitor	Pros	Cons
<p><b>Lalia Learn</b></p> 	<ol style="list-style-type: none"> <li>1. Speech-to-Text support in Lalia Learn enhances <b>verbal learning</b></li> <li>2. Allows users to select content based on their <b>relevant and personal interests</b>, which increases <b>engagement</b> and <b>motivation</b></li> <li>3. Lalia Learn <b>quizzes</b> users immediately after content consumption, for a <b>learning feedback</b> loop. Users are provided with feedback, <b>assessing their own understanding</b> and comprehension</li> </ol>	<ol style="list-style-type: none"> <li>1. The choice of content is from a <b>fixed pool</b> which <b>limits exposure</b> to a broader range of contexts</li> <li>2. The speech-to-text feature supports verbal learning but <b>lacks a conversational context</b>, which may <b>limit practical language use</b></li> <li>3. Quizzes grade user inputs as <b>right or wrong</b> without providing <b>qualitative feedback</b> or specific areas for improvement</li> </ol>
<p><b>LanguaTalk</b></p> 	<ol style="list-style-type: none"> <li>1. LanguaTalk accepts only the top 10% of tutor applications, boasting <b>top quality coaches</b></li> <li>2. Provides <b>personalised assistance</b> through real coaches</li> <li>3. Features <b>interactive transcripts</b> as users talk and one-click word translations</li> <li>4. Utilises <b>advanced AI</b> for quick practice conversations and role plays</li> </ol>	<ol style="list-style-type: none"> <li>1. Coaching sessions are timed based and <b>limits accessible for users</b> based on tutors availability</li> <li>2. <b>Lack of specific context setting</b> for AI interactions, limiting exposure to the natural variability and challenge of real-world scenarios</li> <li>3. More effective features and personalised learning experiences requires subscription to the pro plan, which could be a <b>financial barrier</b></li> </ol>

<p><b>Lingosnap</b></p> 	<ol style="list-style-type: none"> <li>1. Leverages on imagery learning through the <b>Pictorial Superiority Effect</b> to <b>enhance memory retention</b> and effectiveness</li> <li>2. Allows users to upload their own photos, <b>connecting learning to personal experiences</b> or surroundings, which <b>increases relevance and applicability</b></li> </ol>	<ol style="list-style-type: none"> <li>1. Focuses primarily on <b>building vocabulary</b> through images, which may <b>lack context-based or conversational learning</b></li> <li>2. Heavily dependent on images, which could <b>limit understanding and language use</b> in situations with <b>non-visual aids</b> (such as giving explanations)</li> </ol>
---	---	--

Through analysing the pros and cons of competitors in the language learning market, we distinguish SpeakEasy by building on the benefits that these applications bring, while minimising their shortcomings.

SpeakEasy



SpeakEasy focuses on **enhancing conversational and verbal skills** by integrating **context-dependent scenarios** with imagery, conversational context, and speech-to-text technologies. This holistic approach allows learners to **practise language use in a variety of real-world situations**, fostering comprehensive conversational competence.

SpeakEasy offers **contextual variety** for users to select from pre-set scenarios for language practice. However, users may also upload their own photos to **create personalised learning contexts**. This flexibility ensures that users are not confined to a limited set of situations and also confers the benefits of **Pictorial Superiority Effect**, similar to Lingosnap. Users delve into their customised scenarios through their own uploaded photos allowing them to engage in **deeper and personal conversations**.





SpeakEasy **utilises advanced AI**, similar to LanguaTalk, to provide an **interactive and engaging conversational learning experience**. Users benefit from a dynamic call-and-response format that simulates real-world interactions, enhancing the practical application of language skills in everyday scenarios.

Language is about effectively conveying ideas, and not about right or wrong answers. SpeakEasy provides **qualitative explanations, corrections, and suggestions for improvement**, empowering learners to communicate more effectively. This enhances the learning feedback loop, inspired by Lalia Learn.

SpeakEasy is accessible 24/7, enabling learners to practise language skills **at their own pace and on their own schedule**. This flexibility ensures learners have the opportunity to learn and improve without constraints.



## Features Comparison

	Lalia Learn	LanguaTalk	Lingosnap	SpeakEasy
				
Supports conversational skills	✓	✓		✓
Personalised context	✓		✓	✓
Quantitative Feedback	✓			✓
Qualitative Feedback		✓		✓
Transcript Support	✓	✓		✓
Utilisation of Advanced AI		✓	✓	✓
Imagery Learning			✓	✓
24/7 Accessible	✓		✓	✓

## Milestone 2

SpeakEasy is a language learning platform designed to immerse users in **real-world scenarios**. By leveraging generative AI, the app creates **dynamic, personalised** conversations that adapt to each user's skill level, interests, and learning pace. SpeakEasy aims to provide a more natural, engaging, and context-rich language learning experience, allowing users to practise speaking in various situations without fear of judgement.

### Objectives

1. **Provide Contextual Language Learning:** Immerse users in realistic scenarios to help them learn vocabulary and grammar in context, making it easier to apply their knowledge in practical situations.
2. **Offer Personalized and Adaptive Learning:** Utilise generative AI to create dynamic conversations that adapt to each user's skill level, interests, and pace, providing a customised learning experience.
3. **Enhance User Engagement Through Dynamic Content:** Offer varied, AI-driven scenarios and content to prevent repetitive and monotonous learning.
4. **Provide Meaningful Feedback:** Offer personalised feedback to help users identify areas for improvement.

### User Stories

1. As an intermediate language learner (of Mandarin), I want to practise speaking the language regularly to **maintain my fluency** and prevent losing proficiency.
2. As an intermediate language learner (of Mandarin), I want to practise **common real life scenarios** so that I am prepared to face them in real life.
3. As an intermediate language learner (of Mandarin), I want to **receive feedback** so that I know where to improve.
4. As an intermediate language learner (of Mandarin), I want to **explore a variety of scenarios** so that I stay engaged and motivated to learn.

### Additional Notes

- For our MVP, we decided to focus on building product capabilities for only **one language**. As such we chose Mandarin, as all of our group members are **familiar** with that language.
- We are targeting **intermediate** language learners and not complete beginners for this application. We assume the user should already have **foundational knowledge** of the language like understanding of basic words but want to go **beyond their current level** of understanding, to a more **conversational level** of mastery of the language. We will elaborate more in Milestone 4.

## Milestone 3

### 1. Dynamic Content to Maximise Engagement

SpeakEasy's scenario-based context learning is more than just a novel feature - it is a **dynamic engagement tool** that evolves with cultural and global **trends**. By **integrating relevant events**, such as the Olympics, into our learning modules, we ensure that our content remains **current and engaging**. This **continuous refreshment** of scenarios, driven by user data and global trends, allows us to offer **uniquely relevant and enjoyable** learning experiences.



#### Sailing towards Olympic Success

Discuss with a friend your enjoyment of the recent Olympics, focusing on Singapore's participation and the significance of the country



#### F1 Racing

Invite your friend to attend the upcoming F1 race in Singapore together. Share your excitement about the vibrant atmosphere, the illuminated

*(Example image of Olympic and F1 Racing scenarios)*

As users interact with the platform, we may also collect data on the **popularity, user engagement and learning effectiveness** for various scenarios. This data informs our content strategy, allowing us to identify and **transform frequently customised scenarios into fixed offerings**, for better **accessibility and relevant** learning experience for the wider user base.

These strategies ensure that our content **remains on the cutting edge** of both **linguistic relevance** and **user preference**. Hence, allowing us to keep a **competitive edge** through maintaining relevant and popular scenarios.

## 2. Personalised Feedback System

Generative AI is utilised to provide in-depth, contextual feedback on user inputs. This goes beyond correct or incorrect responses to include **qualitative explanations** on usage, suggestions for alternative expressions, and guidance on cultural nuances. In the future as we have AI integration, it is possible to utilise historical user data, for our application to monitor and **adapt to user's performance**. Through this, we can refine our LLM to generate feedback which **reflects their evolving progress**, giving them a better **sense of achievement** and providing users with a more **personalised learning experience** through SpeakEasy. This unique feedback mechanism is possible through users' past data and is hence, **not easily replicated** by competitors. This strategy **enhances personalisation** but also **boosts user retention rates** by providing a **distinctly tailored** educational journey through SpeakEasy.

## 3. First Mover Advantage

As one of the first platforms to integrate **real-world contexts** with **AI-supported conversational** language learning, SpeakEasy **sets a new standard** for what users expect in language learning applications. Leveraging on the first-mover advantage, we aim to attract and **expand our user base rapidly**. A larger user base allows us to **gather extensive data**, which fuels the **efficient training and refinement** of our LLM. This allows for our application to continually enhance its **conversational capabilities as well as the personalisation aspect** of the learning experience. By **setting a new standard**, and **persistently elevating** these standards based on user data and feedback, SpeakEasy maintains a **competitive edge**, making it challenging for newcomers and competitors to sway our user base, while we stay ahead of the market.

## Milestone 4

The target users of SpeakEasy can be largely divided into the following 3 groups, which are categorised by their motivations to learn a new language - for personal interest, travelling or business.

User Group	User Acquisition Plan
<p><b><u>1. Language Enthusiasts</u></b></p> <p><b>Characteristics:</b></p> <ul style="list-style-type: none"><li>• These users are passionate about <b>learning multiple languages</b> and often seek <b>advanced or nuanced</b> ways to improve their conversational skills</li><li>• They might already be <b>familiar with traditional platforms</b> and are looking for innovative, more <b>engaging</b> tools like SpeakEasy to complement their learning</li></ul> <p><b>Needs:</b></p> <ul style="list-style-type: none"><li>• Contextual, <b>real-life scenario-based</b> language practice</li><li>• Advanced <b>conversational</b> capabilities</li><li>• <b>Personalised</b> learning paths</li></ul>	<p><b>Content Marketing and Blogging:</b></p> <ul style="list-style-type: none"><li>• Develop content on language learning strategies, advanced tips, and the value of <b>contextual learning</b> to attract enthusiasts who are constantly looking to improve</li><li>• Publish articles on <b>language learning blogs</b> or collaborate with polyglot influencers</li></ul> <p><b>Community Engagement:</b></p> <ul style="list-style-type: none"><li>• Sponsor language meetups or polyglot gatherings</li><li>• Engage with forums like Reddit (<a href="https://www.reddit.com/r/languagelearning">/r/languagelearning</a>) or Duolingo's forums to discuss personalised learning.</li></ul>
<p><b><u>2. Casual Learners or Travellers</u></b></p> <p><b>Characteristics:</b></p> <ul style="list-style-type: none"><li>• Individuals looking to <b>improve their existing knowledge</b> of a language they picked up, for <b>travel</b> or general interest.</li><li>• These users often seek <b>practical, conversational language skills</b> for specific scenarios, such as ordering food, asking for directions, or basic social interactions.</li></ul>	<p><b>Targeted Social Media Ads:</b></p> <ul style="list-style-type: none"><li>• Create <b>fun and engaging ads</b> on platforms like Instagram and TikTok, showcasing SpeakEasy's <b>practical applications</b> for travel and general scenarios</li></ul> <p><b>Collaborate with Travel Bloggers/ Vloggers:</b></p> <ul style="list-style-type: none"><li>• Partner with <b>travel influencers</b> to promote SpeakEasy as a helpful tool for learning the basics before a trip</li></ul>

<p><b>Needs:</b></p> <ul style="list-style-type: none"> <li>• Simple, <b>scenario-specific</b> language lessons</li> <li>• <b>Flexibility</b> to learn at their own pace</li> <li>• <b>Practical</b> conversation practice.</li> </ul>	<ul style="list-style-type: none"> <li>• Collaboration can be extended to <b>travel sites</b> like Airbnb, Trivago and airlines to cross sell SpeakEasy service</li> </ul>
<p><b><u>3. Busy Professionals</u></b></p> <p><b>Characteristics:</b></p> <ul style="list-style-type: none"> <li>• Professionals who need to improve language skills for <b>work, networking, or international business</b> interactions.</li> <li>• They often have limited time but need to <b>master situational conversations</b> relevant to their careers, such as meetings, negotiations, or presentations.</li> </ul> <p><b>Needs:</b></p> <ul style="list-style-type: none"> <li>• Flexible, on-demand learning <b>tailored to business or professional scenarios</b>, including real-time conversational practice.</li> </ul>	<p><b>LinkedIn Ads:</b></p> <ul style="list-style-type: none"> <li>• Use <b>LinkedIn</b> to <b>target professionals</b> in multinational companies or industries requiring multilingual communication</li> </ul> <p><b>Corporate Partnerships:</b></p> <ul style="list-style-type: none"> <li>• Partner with <b>businesses</b> offering SpeakEasy as part of their employee development or <b>language training programmes</b></li> </ul> <p><b>Webinars:</b></p> <ul style="list-style-type: none"> <li>• Host webinars on "How to Improve Your Business Communication Skills" or "Multilingual Networking Skills"</li> <li>• While emphasising SpeakEasy's personalised approach to <b>professional language learning</b>.</li> </ul>

# Milestone 5

## Current Features

### 1. Conversational learning experience through speaking with Gen AI

We think that this is the foundational feature of SpeakEasy as it creates dynamic and interactive conversations that **mimic real-life scenarios**. Since SpeakEasy's main goal is to help users improve their **conversational fluency**, this feature directly addresses that need. This is SpeakEasy's **primary appeal**, offering a realistic and engaging, but safe environment for practice and learning.

### 2. Users create personalised scenario with their own context and image

By allowing users to customise their **bespoke scenarios**, SpeakEasy can effectively **address the key pain points** in language learning, i.e. the boring, repetitive, irrelevant contents and be more meaningful to users by providing them **exactly what they want**. This feature greatly enhances users' language learning journey, giving SpeakEasy an **edge** over our competitors. Through SpeakEasy, users form a **stronger connection with the language** that they are learning through these personalised scenarios. This feature increases **user satisfaction** and is crucial in user retention, aiding in the long-term **success** of SpeakEasy.

### 3. Context Refinement

While users create custom scenarios, our application sanitises and refines users' initial context as **clear and actionable objectives**. This ensures users can **meaningfully engage with their customised scenarios**. Integrated guard clauses and advanced image processing enhance the context setting, providing a **richer and tailored learning experience**. Should user inputs be incomprehensible or inappropriate, the system automatically falls back to a default scenario where the user describes their day to a friend so as to **maintain the integrity and safety** of our application.

### 4. Personalised suggestions and feedback on user's input

We believe that suggestions and feedback are crucial for helping users improve their language skills, and missing this feature would make SpeakEasy an incomplete learning app. It is important for language learners to **receive external feedback** (i.e. from SpeakEasy), and by making it **personalised**, SpeakEasy ensures that the learning experience is more **relevant** to each user. Personalised feedback provides **constructive, tailored insights** rather than generic comments, which enhances the overall learning process.

## Future Features

### 1. Versioning for past scenario conversation history

This feature is relevant because it allows users to **revise, reflect on improvements** and see areas where they struggled. Having access to past scenarios can also let users **compare their initial attempts** with improved conversations. However, we think that this is not essential to the core purpose of SpeakEasy which is to provide immediate and real-time conversation practice. Versioning does not contribute directly to real-time language practices hence is **less urgent** for the initial release.

### 2. More scenarios, more languages

Offering a **wider variety of fixed scenarios** and language support would significantly expand SpeakEasy's appeal to a broader audience and enrich users' experience. While it enhances the app's versatility, SpeakEasy's priority on providing a **high-quality learning** experience is not affected and missing these will not affect the core conversational experience that SpeakEasy provides. Also, this feature is resource-intensive as we need extensive content creation, and is best left for later development phases **after the foundational features are fully implemented**.

### 3. Video-based scenarios

By allowing users to create their own **scenarios in videos**, SpeakEasy can provide a more **immersive experience** because videos enhance the **realism and depth of the context** and are more engaging. We also believe that dynamic scenarios can help users learn more effectively. However, since this feature is complex and requires more time and effort to implement, we decided to keep it as a future feature.

### 4. Gamification

By **gamifying** the language learning experience, SpeakEasy **motivates** users to stay **committed** to their learning journey by driving engagements and objectives. It can also help in keeping users continuing using our app for a longer time. However, this does not directly contribute to the main goal of SpeakEasy in helping users **gain conversational skills** to prepare them for the real-world. Hence, it is not necessary for the app's initial release.

### 5. Group conversations (e.g. a 3-way conversation with friends and an AI)

This feature allows multi-person interaction, in Zoom-like conversations where users can invite friends to join in their SpeakEasy conversations. Supported by video streaming and AI responses, this feature would **encourage collaborative learning**, a highly engaging way for users to collectively practise their language skills. It could also help **expand our**



**user base**, when users invite their friends to join the app. While it holds value for group learning, it is not essential for the MVP. Group conversations **introduce additional complexities** to the application and **require more time and resources to develop**, which would be more suitable in later-stage development.

## Milestone 6

### Pricing Strategy and Features of each Tier

We will use a freemium model with tiered subscriptions. The few tiers will be:

1. **Free Tier:** \$0/month
  - Limited access to basic scenarios and two scenario generations per day.
2. **Standard Tier:** \$9.99/month
  - Provides access to a wider range of scenarios, 10 scenario generations per day and more frequent practice sessions.
3. **Premium Tier:** \$19.99/month
  - Offers unlimited access to all scenarios and scenario generations, priority for new content and languages, and advanced features like pronunciation analysis.
4. **Enterprise Tier:** Custom pricing for educational institutions (estimate \$29.99/user)
  - Tailored for schools or companies, with all features from premium tier and additional features like progress tracking and custom scenario creation.

### Rationale For Freemium Model and Subscription Model

Firstly, the freemium model creates a **low barrier to entry** for users. The free tier allows potential users to try out the platform without any financial commitment. This is crucial in the competitive language learning market where users often try multiple apps before settling on one. The free tier serves as a showcase for SpeakEasy's unique selling points (USP). By offering a small bite of SpeakEasy, we can attract users who might be hesitant to pay upfront for an unfamiliar concept.

Secondly, while users interact with our platform, we can upsell and **persuade them to purchase the paid features**. As users engage with the free content, they naturally encounter the limitations of the free tier. This creates organic opportunities to highlight the benefits of paid tiers. For example, a user might reach a daily limit on scenario generations and if they want to continue learning, they might consider upgrading.

Lastly, the freemium model encourages more users to be on the platform, allowing us to **collect data and feedback to improve** on SpeakEasy. More users provide valuable data on usage patterns, preferred scenarios, common generated scenarios and more. This data can be used to improve the AI models, refine the learning experience, and inform decisions about which features to develop or prioritise in paid tiers.

We chose the subscription model because the revenue will be used to **sustain the scenario creation** by design professionals. In the future, we will also have the revenue to possibly develop and train our custom AI models. We delve into the breakdown of costs below.

## Revenue Streams

1. **Subscription Fees:** The primary revenue stream, providing recurring income on a consistent basis.
2. **Enterprise Contracts:** Potentially significant revenue from educational institutions or corporate clients.
3. **Partnerships:** Revenue sharing with content creators, Youtubers or language experts who contribute to scenario development.

## Factors Influencing Pricing Decisions

We believe our SpeakEasy is of **high perceived value**. The immersive experience that feels more like interacting with a real person than using traditional language learning tools or another application. This personalised approach creates a higher perceived value for users. As such, we are positioning ourselves as a premium product but **keeping prices comparable** to other well-known language apps such as DuoLingo which is priced at roughly \$10/month. Another competitor, LanguaTalk, charges \$11 to \$90 per hour depending on the coach. In comparison, we position ourselves at a **fraction of what LanguaTalk charges** but provides similar value in terms of language learning capabilities and feedback.

The production costs for our AI-driven language learning platform are not exceedingly high but increases proportionally with the number of users. We are using the model GPT-4o mini which costs \$5.00 per 1 Million input tokens. If we estimate 5000 tokens used per user per day, this \$5 would allow us to serve up to 200 active users per day. The per 30-day month cost would then be \$150 for 200 monthly active users (MAUs). Factoring deployment and hosting costs on Vercel Premium (\$20), Netlify Premium (\$19), Atlas (\$57), our total operating cost for technical infrastructure is about **\$96 per month**. For labour cost, assuming we have 4 part-time software engineers who draw \$400 per month (assumption: \$1200 full time pay, part time work for 1/3 of the time), our total labour cost is **\$1600 per month**. Considering that all users are subscribed to the standard tier, our monthly revenue would be \$1998 (200 x \$9.99) and a **17.8% profit margin** (though realistically there are other cost factors such as office and marketing that have not been considered). Nonetheless, the \$9.99 price point is reasonably justified. It covers our operational expenses (such as LLM, hosting platform or cloud storage) and allows for reinvestment in the platform while providing users with a valuable learning experience at a competitive price.

Looking into the future, developing, maintaining and improving the platform, and possibly adding our own AI models into the mix, would require significant investment in both technology and skilled engineers in the thousands of dollars. These models need regular updates to improve performance and adapt to user feedback. Creating diverse and engaging scenarios for immersive learning demands continuous effort from content creators and language experts. Overall, the cost is very high. The tiered pricing strategy implemented will help distribute these expenses across our user base. Premium tiers will contribute a larger share towards covering these costs. This approach allows us to sustain the high quality of your service while keeping the entry-level tier affordable, thus balancing financial sustainability with market accessibility.

## Milestone 7

LLMs are particularly useful for SpeakEasy as we aim to make language practise more **dynamic and engaging**.

### Contextual Understanding

As users interact with the SpeakEasy using real-world scenarios, the LLMs are able to **retain and process context**, allowing for relevant responses that feel **fluid and coherent**. For instance, when a user uploads a photo and starts a conversation based on the given context, the LLM can dynamically **adjust the conversation to the scenario** and provide **context-aware vocabulary**. By incorporating this contextual information, LLMs can generate conversions with a **sense of immersion** and continuity in language learning, which helps **overcome the "boring repetition"** problem in traditional methods and keeps users interested and motivated.

Additionally, because of LLM's contextual sensitivity, it is able to provide **nuanced, qualitative feedback on user responses** to each scenario. Rather than simply marking responses as right or wrong, LLMs can offer suggestions for improvement, detailed explanations, and alternative replies. This aligns with our goal of helping users learn to **communicate more effectively** in real-world scenarios, rather than focusing solely on technical correctness. The depth of feedback provided by LLMs creates a more **holistic learning experience**, ensuring learners understand not only what to say but also why certain phrases are appropriate in different scenarios.

### Adaptability

Furthermore, because of LLMs' adaptability to **handle a limitless variety** of scenarios, LLMs help SpeakEasy ensure that every user can practise language in a context **relevant to their needs**. Whether for work, travel, or casual conversation, LLMs allow SpeakEasy to provide **scalable, personalised learning** without the need for manual content creation.

### Multilingual Capability

LLMs can support a **broad spectrum of languages**, from widely spoken ones like English, Spanish, and Mandarin to less common languages like Swahili or Finnish. This inclusivity ensures that users from different linguistic backgrounds can benefit from the application, making language learning **accessible to a larger audience**.

## Milestone 8

### Example 1: Prompt for AI Response to Users

```
""
You are a helpful language learning assistant.

User Input: "{user_input}"
Context: "{context}"

Based on the user input and context, provide a meaningful and contextually
appropriate response in Chinese.
If the user's input is irrelevant to the context or in English, you should reproach
them and remind them to continue the conversation in Chinese, focusing on maintaining
relevance to the topic.

If the context objective is not met, encourage the user by concluding your response
with a guiding question.
If the context objective is met, conclude appropriately without further questions.

Response should not be longer than 2 sentences.
""
```

Legend: [ ] are used to capture common named techniques of prompt engineering

1. Using **clear and concise language minimises misinterpretations** by the model. We had intentionally provided **explicit instructions** at the beginning of the prompt to **reduce ambiguity** and enable the model to **execute the task accurately**.
2. **[Context Expansion]** By **embedding context directly** in the prompt and specifying the role as a "helpful language learning assistant," we ensure that the LLM's responses are **relevant** and also **tailored to enhancing** the user's language learning experience. This approach allows the model to generate **constructive feedback**, emphasising on the **educational value** of its interaction with the user.

As the model operates in a **stateless** manner, each session is processed independently. We decided to **incorporate all previous user inputs**, but focusing feedback **exclusively on the latest input**, the model remains **informed about the user's progress** without retaining past state.

3. **[Task Specification]** The prompt is designed to **check if the context objective is met** and to **encourage further conversation** where necessary. This dual approach ensures that the AI not only **assesses the user's response** for accuracy and relevance but also keeps the

dialogue moving forward by asking **guiding questions**.

4. The design of this prompt includes strategies to **prevent adversarial prompts** that could lead to **irrelevant or harmful content** generation. By structuring the prompt to focus the AI on educational outcomes and contextually appropriate language, we **mitigate the risk of the model generating undesirable content**.
5. The response generated by the AI in the primary language is followed by a **separate process for English translation**, maintaining the **integrity of the response** in both languages. This separation ensures that the **translation** does not **influence the original content generation**, prioritising the **educational quality** of the responses.

## Example 2: Prompt for Initial Message for Context

```
""
You should take on the receiving role of the context, without responding to this
prompt.
The user's objective is '{context}'
Output an initial message to initiate a conversation as the receiving role of the
context in Chinese.

For example, if the user has to order food, you should assume the role of a waiter.
For example, if the user has to make a phone call, you should be the receiver.

It should be short and less than 2 sentences.
""
```

1. **[Few-shot Prompting]** Previously, the model's response was unstable with zero-shot and one-shot prompting, sometimes behaving as the user trying to achieve the context objective. By **including examples** directly in the prompt, the model is **better guided** on how to perform the task accurately. This technique not only **clarifies the expected role** for the AI and **enhances** its ability to generate **relevant and context-specific outputs**.
2. The specification that the initial message should be concise, limited to two sentences, ensures **clarity and digestibility** of the feedback. Without this constraint, the initial message output tends to be **excessively verbose**. This enables the model's output to be **concise** and facilitate a more **efficient and engaging interaction**.
3. The structure of the prompt itself was carefully **designed with clarity** in mind. By incorporating a **newline to distinctly separate different sections** of the prompt, each part is **clearly delineated**. This methodical structuring ensures that the model **comprehends and adheres to each aspect** of the task.

### Example 3: Prompt for Suggested Responses

```
"""
You are a helpful language learning assistant.
Provide a structured list response containing suggestions in responding to the
previous message and strictly nothing else.
Suggestions should be at most 1 sentence.
No ```code declaration needed, just the list object.

Previous Input: "{prev_message}"
Context: "{context}"

Requirements:
- first: A simple and relevant response to the Previous Input in Chinese.
- second: A meaningful and contextually appropriate response to the Previous Input in
Chinese.
- third: A meaningful and contextually relevant response to the Previous Input in
Chinese.

Return a LIST of suggestions, example: [first, second, third]

Ensure the response is in valid list format.
"""
```

1. The **structure of the prompt** is **explicitly designed** to guide the LLM effectively: it starts with a brief **introduction**, proceeds with the **previous input**, **scenario context**, and follows with **output requirements of increasing complexity**. This sequence ensures that the model has a clear direction and purpose throughout its processing, enhancing the relevance and accuracy of its feedback.
2. The requirements of the LIST output are described in a **well-organised** manner and described through with the use of **an example**. This aids the model in **efficiently parsing the multiple outputs**. By **capitalising** the LIST output, the prompt emphasises the specific format needed, which is essential in order to parse this output into JSON effectively.
3. **Specifying the length** of the output is also essential. Initially, we encountered issues where responses were truncated midway due to exceeding the `max_tokens` limit. By specifying the output length, we effectively prevent outputs from being prematurely cut off and receive complete, concise outputs.



## Milestone 9

### Choice of LLM and Provider

Our choice of LLM is GPT-4o-mini provided by OpenAI.

### Considerations

Prior to narrowing down the choices of LLMs and respective models, we identified several crucial factors in which we want our model to behave, to streamline our decision-making process. We narrowed it down to these the following key factors:

- Popularity
- Cost
- Language Support
- Multimodality
- Performance of model

### Popularity

With the rapidly expanding field of generative AI, and numerous LLMs readily available, it is impractical for us to evaluate and consider every model. The popularity of a model/provider was the **initial** factor in our selection process which can be indicative of its **reliability, community support, and the breadth** of its ecosystem. Well-regarded models tend to also be supported by **extensive documentation and user guides** which contribute to smoother development and easier troubleshooting. These attributes of popular LLM providers helped us to **streamline our decision-making process**, allowing us to focus on options that promise a supportive and dynamic environment for deployment.

Rank* (UB)	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
17	<a href="#">Claude 3 Opus</a>	1248	+3/-2	165503	Anthropic	Proprietary	2023/8
32	<a href="#">Llama-3-70b-Instruct</a>	1206	+3/-2	163849	Meta	Llama 3 Community	2023/12
46	<a href="#">Claude 3 Haiku</a>	1179	+3/-2	114780	Anthropic	Proprietary	2023/8
37	<a href="#">Claude 3 Sonnet</a>	1201	+3/-3	113039	Anthropic	Proprietary	2023/8
57	<a href="#">Llama-3-8b-Instruct</a>	1152	+2/-2	109335	Meta	Llama 3 Community	2023/3
17	<a href="#">GPT-4-1106-preview</a>	1251	+2/-3	97378	OpenAI	Proprietary	2023/4
14	<a href="#">GPT-4-Turbo-2024-04-09</a>	1257	+3/-2	92973	OpenAI	Proprietary	2023/12
52	<a href="#">GPT-4-0613</a>	1163	+3/-2	91629	OpenAI	Proprietary	2021/9
6	<a href="#">GPT-4o-2024-05-13</a>	1285	+3/-3	90695	OpenAI	Proprietary	2023/10
17	<a href="#">GPT-4-0125-preview</a>	1245	+4/-2	90560	OpenAI	Proprietary	2023/12
42	<a href="#">Command R+ (04-2024)</a>	1190	+3/-3	80852	Cohere	CC-BY-NC-4.0	2024/3
12	<a href="#">Gemini-1.5-Pro-001</a>	1259	+3/-3	80656	Google	Proprietary	2023/11



(Most popular LLMs, among 145 models and 1.8M+ votes

Source: <https://lmarena.ai/?leaderboard>)


From here, we were able to narrow down to the more popular LLMs, Claude-3, Llama, GPT-4 (and GPT-4o), Cohere and Gemini.

Cost

Cost is a critical factor in the selection of an LLM provider because it directly impacts the project's **budget and scalability** of our project. Keeping costs affordable is essential for enabling comprehensive testing phases, iterative enhancements, and ensuring the widespread deployment of the application. As we are responsible for the costs of the API, it was **essential to maintain a realistic budget**. Maintaining affordability also enables us to sustainably **experiment and explore various prompting techniques** to achieve the best results. This approach ensures that our project remains **financially viable** while allowing for **continuous optimisation and improvement** of our application.

	<div>Claude-3</div> <div></div>	<div>Llama-3</div> <div></div>	<div>Command R+</div> <div>Cohere</div> <div></div>	<div>Gemini</div> <div></div>
Price/1M Input Tokens	\$15	\$2.65	\$2.50	\$0.075
Price/1M Output Tokens	\$75	\$3.50	\$10	\$0.30





When comparing different GPT models, finding the right **balance between cost and performance capabilities** is also crucial. Earlier models like GPT-3 are less expensive but did not have multimodal capabilities or improved context understanding. On the other hand, heavier and advanced models such as GPT-4, are **costly but provide substantial improvements** in linguistic accuracy, context understanding, and output generation.

	<div>gpt-3.5-turbo</div>	<div>gpt-4-turbo</div>	<div>gpt-4o</div>	<div>gpt-4o mini</div>
Price/1M Input Tokens	\$0.50	\$10	\$5	\$0.15
Price/1M Output Tokens	\$1.50	\$30	\$15	\$0.60

Through our analysis, we excluded Claude-3 due to its higher pricing tier. We continued to consider the suite of GPT models, as they were **easily interchangeable** and offered flexible options for our needs.

### Language Support

Language support is essential in meeting the **diverse linguistic demands** inherent in our language learning application. We require a model proficient in multiple languages, including English and Chinese, and is sensitive to cultural nuances. This comprehensive language support ensures our application can deliver high-quality responses across various languages as we scale. The output of the LLM directly influences the user's learning experience. Hence, the ability of the LLM to accurately reflect linguistic and cultural contexts is a crucial metric in the evaluation of these potential models.

	GPT	Llama-3	Command R+	Gemini
				
Language Support	More than 80 languages	8 Languages	109 Languages	More than 40 languages
Source	<a href="#">Link</a>	<a href="#">Github</a> , <a href="#">Link</a>	<a href="#">Link</a>	<a href="#">Link</a>

Among the eight languages supported by Llama-3, Mandarin Chinese, which is the primary language for this MVP, is **not included**. The **limited language support** of Llama-3 also restricts the **scalability of our application** to other languages, hence further solidifying our decision to exclude it from our selection.

### Multimodality

The primary function of our chosen LLM is to generate and facilitate conversational interactions in the target learning language. Additionally, we utilise the LLM for the processing of customised scenarios, where a **multimodal approach** is beneficial. This allows for the integration of user-uploaded images to establish relevant contexts for these scenarios, enhancing the overall learning experience by setting more accurate contexts. Consequently, we excluded Command R+, which is **only limited to conversational capabilities**. Within the GPT suite, we opted for gpt-4o and gpt-4o mini, as these models support multimodal inputs, aligning with our application's dual-functionality needs.

### Performance of Model

The performance of our chosen LLM critically impacts the user experience and the overall effectiveness of our language learning application. These advancements can significantly decrease

the need for user corrections and follow-ups, thereby reducing long-term operational costs and improving the overall user experience. The selection process is crucial to ensure that the chosen model **aligns with our developmental needs**.

Key performance metrics include **accuracy**, **response time**, and the model's ability to handle **nuanced language tasks**. Our evaluation employed specific user prompts to assess these metrics. Comparing GPT-4o-mini and Gemini-1.5-flash, GPT-4o-mini gave slightly faster, and more thought-provoking results. (Refer to [Annex](#) for results)

In our evaluation, we also observed that the conversational responses generated by the GPT models were **notably more creative**, leading to **dynamic and engaging interactions** with users. Our findings align with external comparisons detailed in LLM-related blogs, reinforcing our results.

Performance	Excellent for creative text generation, good for some factual tasks	Strong overall performance across various NLP tasks	Focus on safety and factual accuracy can impact creative freedom	Performance details limited, but likely good across various tasks
Accuracy	Good for creative text formats, factual accuracy can be a concern	Generally high accuracy across various tasks	Focuses on factual accuracy and minimizing bias	Good across various tasks, potentially strong on factual accuracy

(Source: <https://geniusee.com/single-blog/llm-model-comparison-which-one-to-use>)

This contrasted with the outputs from the Gemini model, which, while stable and relevant, sometimes lacked variety and engaging qualities seen in GPT’s responses. The “creativity” of the LLM is an essential factor in **sustaining user interest and improving language learning outcomes**.

### Conclusion

We opted to integrate the OpenAI suite of GPT models into our application which produced **engaging and relevant** outputs. The **easy interchangeability** between GPT models was a significant advantage which allowed for flexibility; We could test initially with lower-cost options and then seamlessly transition to more robust models for deployment. OpenAI's models are also **developer-friendly**, featuring **comprehensive documentation** and **robust community support**, which streamlined our development process. Additionally, their support for **multimodal inputs** and **extensive language options** aligns perfectly with the needs of our language learning application, making them an ideal choice for enhancing user experience and application functionality.

For our MVP, we selected GPT-4o-mini for its optimal balance in **cost-effectiveness** over GPT-4o, while still retaining multimodal capabilities, and quick response speed. This model is also resource-efficient, enabling it to handle our prompts with both high accuracy and ease.

## Model Setting

Adjusting these model parameters allows us to finely tune the AI's responses to align with our goals for user engagement and educational value.

### Max Tokens

The maximum number of tokens set for generating responses is set at 1000, to **control usage effectively**. The actual **constraint on the length of outputs** is managed through specific instructions in the prompt text, as outlined in milestone 8.

### Temperature

For conversational modules that require more creative and varied responses, we adjust the temperature to be **higher, at 0.8**. This setting allows the AI to generate more diverse and imaginative replies, enhancing the **engagement and stimulation** of the conversation. A higher temperature allows our model to explore a **wider range of cultural contexts**, phrasings and concepts, making interactions more dynamic and less predictable.

Conversely, for functions like context refinement and generating the initial message for custom scenarios, we set temperature to be **lower at 0.2**. This consistency is crucial as these prompts are generated **only once** during scenario creation. A lower temperature ensures a **stable context and initial message**, facilitating seamless scenario creation for users.

### Frequency/Presence Penalties

Presence penalties encourage the model to use different words, while frequency penalties prevent the model from using the same words repeatedly. ([source](#))

In the context of our language learning application, we strategically employ presence penalties to **diversify the vocabulary exposed** to users. We have chosen not to implement frequency penalties, as certain niche learning scenarios may require **repeated reference to specific objects or concepts**, making it necessary to use repeated terms.

For conversational modules, we set a **presence penalty of 0.3**, and encourage a more varied vocabulary in the LLM responses to users. This ensures learners are exposed to a wider range of vocabulary, enhancing their learning experience while interacting with the application.

For customising scenario functions like context refinement and the generation of initial messages, presence and frequency penalties are **not employed**. In these cases, **consistency and clarity are prioritised over lexical variety**, ensuring that users have a stable and clear basis while customising their learning scenarios.



## Milestone 10

### Product Name

Our product name is SpeakEasy, it plays on the idea of "speakeasies" from the Prohibition era, which were secret bars where people could freely converse and socialise. The name suggests a **relaxed, informal atmosphere** where language learners can speak easily, implying that the app makes **learning to converse in a new language feel natural and effortless for our users**.

Another alternative was FluentFlow but we felt that it could not capture the idea of learning a new language as well as SpeakEasy. On the other hand, SpeakEasy is a clever wordplay that also makes it stand out more than FluentFlow, and could also add a sense of **fun and approachability** to the learning process. "Speak" and "Easy" are **simple English words** that are widely understood, as such someone who does not speak English would still be able to recall the name easily, appealing to a global audience of learners. However, "Fluent" and "Flow" are more obscure words. Given that our application can be used by everyone in the world, many of whom don't speak English, SpeakEasy would be a more optimal choice as it is easier to understand and remember, which encourages people to use the application.

### Logo

We chose to use the pinkish-red as the primary colour as it is often associated with energy, **enthusiasm, and passion**, reflecting the active, engaging nature of learning a new language on our application. Purple colour is associated with **wisdom**, signalling that the app will help users unlock new knowledge through learning a new language on SpeakEasy. The gradient itself represents **growth or progress**, implying that users will evolve from beginners to confident speakers as they use the app.



The two overlapping speech bubbles reinforce the **idea of conversation**, which is at the heart of the app. The speech bubble on the right is coming from above symbolising the concept of receiving guidance. Our learners receive this helping hand during the learning process from our AI when they practise in real time.

## Tagline

Our tagline "Speak Freely" suggests that our application creates an environment where users can **practise without fear of judgement or making mistakes**, which is an ideal environment for language learners who may feel self-conscious when practising a new language. On top of that, "Speak Freely" also rolls off the tongue nicely when said together with SpeakEasy ("SpeakEasy Speak Freely!"), it is very **catchy** and **easy for users to remember**.



# Milestone 11

## Frontend/UI

Since we are building a Single-Page application, we considered **React**, **Vue** and **Angular** because they are optimised for building dynamic user interface and client-side routing. They also have a large community and have a lot of support for third-party libraries.

We chose **React** because of the **flexibility** it provides. React is a library, not a full-fledged framework like Angular. This gives us the flexibility to structure our app however we want. Angular, being a framework, enforces strict structures and predefined methods (e.g., RxJS, Dependency Injection) and Vue leans more toward a comprehensive framework like Angular, which provides more built-in tools than React.

Another reason we chose React is because all the group members are most **familiar** with it compared to Vue and Angular, ensuring that we all can hit the ground running with frontend development instead of having to spend additional time learning a new framework.

We also decided to use **Typescript** instead of **Javascript** as the language because **TypeScript** ensures type safety which helps in easier and early bug detection. We think it is crucial as for a project with limited time, we do not want to spend too much time debugging.

When choosing bundler for our React application, we chose **Vite** over **Webpack** because it provides a quicker development server and updates changes quicker during the development process with Hot Module Replacement (HMR). For styling, we decided to use **Tailwind CSS + shadcn** as one of our team members has used them before and it provides the benefits of using pre-built components and styling flexibility.

## Backend

We use **Django** for our backend because everyone is already familiar with the framework compared to alternatives like **Express** or **Flask** which some of us do not have experience in. Hence, we decided to go with Django to eliminate the overhead of learning a new framework and also, everyone can work on both the frontend and backend in parallel to reduce communication overhead and speed up development process within the limited time.

Also, since Flask and Express are minimalistic, we imagine that we would require adding packages or extensions during the development process and some of us who are not familiar with the frameworks might run into issues, so we decided not to use them.

## Database

For storing our app-related data, we decided to use **NoSQL** because NoSQL makes it flexible if we want to include more complex unstructured data in the future (eg. conversation history). For **relational databases**, they are more troublesome for rapid prototyping since we need to have a fixed schema and every little change to the schema would require more effort and time to ensure that there is no error. Also, it is cumbersome for our app to scale as we add more features and store more unstructured data.

Additionally, since NoSQL databases are designed to have **high availability**, we think it's a good fit for our app as we want our application to be **readily available for users at any point in time**. On the other hand, relational databases prioritise consistency and will sometimes lead to reduced availability due to updating of data, which is not what we want as users may be greatly turned off by slow loading applications.

We also decided not to use **vector databases** or **graph databases** because our application does not fit their use cases and we do not benefit from using them.

We then decided to use **MongoDB** instead of **Firebase** to store our app data. While both databases have similar pricing, MongoDB has better querying and advanced search capabilities compared to Firebase so we think it is more suitable to store app data with **MongoDB**.

However, since **Firebase** has built-in user authentication service with providers like Google and is easy to set up, we decided to use Firebase for user authentication so that we do not have to deal with that.

We used **Google Cloud Storage (GCS)** as our image store instead of storing the image data as a binary large object in MongoDB. GCS is specifically used for file and object storage data (like images, videos, backups, documents, etc.) at scale. On the other hand, MongoDB, while it can handle large files by breaking them into smaller chunks, is primarily a database meant for structured and semi-structured data, not for efficiently storing and delivering large binary objects.

## Testing

We used **Postman** for testing our APIs during the development process. Postman provides an intuitive and easy-to-use interface that allows us to send GET, POST, PUT, DELETE, and other HTTP requests with just a few clicks. It also allows us to view the request-response cycle, including headers, body, status code, which will be good for debugging our APIs

Link to the Postman Folder:

<https://speeding-space-551894.postman.co/workspace/3216-assignment3~4c873525-a14c-4f8a-94c2-e7300944e0e9/collection/16603346-256f1d84-6f42-4990-a660-068f0408e130?action=share&creator=16603346>

## Hosting

For the frontend, we decided to host it on **Netlify** because it is easy to set up the CI/CD pipeline and for the deployment process. Even though **Vercel** provides the same benefits, Netlify is more optimised for hosting Single-Page Applications generated by client-side rendering frameworks like React as compared to Vercel which is optimised for server-side rendering.

We have also considered **Github Pages** but it does not support environment variables which our app needs, so we decided not to use it for frontend hosting.

For backend hosting, we decided to use **Vercel** to host because it is simple to use and integrates with Github and we are familiar with using Vercel. We have also considered other popular platforms like **Heroku** but it is no longer free to use, so we decided not to use it. Other cloud platforms like **AWS/Google Cloud/Azure** are too complex for our use case due to the large number of configurations, hence we use Vercel which is simpler and requires minimal configurations.

## Authentication

We decided to use third-party services to do authentication because we are not familiar with the security side and developing our own algorithms is **not as safe as what Google can provide** and will take a huge chunk of our development time. We decided to use **GoogleOAuth** since Google has a wide range of users spanning different backgrounds. Implementing GoogleOAuth is also easy because Firebase has the built-in service to handle GoogleOAuth. Also, it allows users to sign up without having to create a separate account for our app and hence streamlines the sign up/login process. We decided not to use **FacebookOAuth** and **GithubOAuth** even though Firebase supports them because their user base is not as big as Google's.

## Milestone 12

### Workflow 1: Interactive Conversation with AI assistant

#### Workflow 1.1: Speech to Text (User Inputs):

Users speak directly to the AI, which converts their speech into text, which is a natural mode of interaction to mimic real-world language use.

#### Workflow 1.2: LLM generates response to user inputs and text-to speech

Our LLM analyses users' input and generates an appropriate response, also with regard to the context. This response is converted back to speech, facilitating a smooth conversational flow.

#### Alternative: Text-Only Interaction

Users type their responses in text format, and the AI provides direct responses also in text format, focusing on reading and writing skills.

#### Conclusion and Rationale

Our approach enhances conversational skills by mimicking real-life interactions, making the learning process **realistic, engaging and dynamic**. As our application is **targeted for intermediate learners** with some language experience, this conversational mode facilitates **auditory learning**, which is crucial for **pronunciation and listening skills**. A text-only interaction model **offers limited differentiation** from existing solutions like software translators or direct language learning model (LLM) chatbots. The interactive conversation also allows learners to be assessed on their coherence of language, grammar and vocabulary.

### Workflow 2: Selecting Scenario to Practice

#### Workflow 2.1 (Alternative 1): Fixed Scenarios only

The application uses a fixed set of predetermined scenarios that are **designed to be universally relevant** and **reflect day-to-day contexts**. This approach ensures that each scenario is **realistic**, for **practical and structured** learning experiences that are consistent across all users.

**Limitations:** While this method offers controlled learning environments and predictable outcomes, it may not fully **cater to individual users' specific interests, cultural backgrounds, or unique learning needs**.

## Workflow 2.2 (Alternative 2): Customised Scenarios only

Users have the ability to upload their own images, and create personalised learning scenarios. This customisation allows users to tailor the context and dialogue based on visual cues from their uploaded content, making the learning experience **highly relevant and engaging**.

**Limitations:** Users are forced to come up with their own scenarios which they may **spend excessive time crafting, diverting their focus** from practising and learning the new language.

## Conclusion and Rationale

Recognising the strengths and limitations of both fixed and customised scenarios, we decided to have the best of both worlds by taking a hybrid approach. This strategy enables us to provide a **reliable foundation for structured learning** through fixed scenarios while offering the **flexibility of customization** for those who seek a more personalised experience.

By integrating both fixed and customisable scenarios, our application adapts to the needs of a wide variety of learners. This hybrid approach positions our application as a **versatile and attractive option** in the competitive language learning market, striking an **effective balance between consistency and innovation**.

## Workflow 3: Receiving Feedback on Performance

### Workflow 3.1: End-of-Session Feedback

When users end the conversation, the overall feedback for the conversation is consolidated and presented at the end of each learning session. Our application assesses the user's performance based on various linguistic metrics, such as relevance, coherence, and complexity. Users receive a **score alongside qualitative feedback**, offering an overview of their performance and areas for improvement.

Our application also offers detailed feedback **for each individual input** within the session. This feature allows users to **drill down into specific responses** where they might need more focused guidance.

### Alternative 3.1: Continuous Feedback Updates

Instead of receiving feedback aggregate at the end of the conversation, an alternative considered was **providing real-time feedback** after each input.

## Workflow 3.2: Hints for Enhancing User Engagement

If the user encounters difficulties or pauses during interaction, our generative model also provides contextually appropriate **suggested responses in increasing complexity**. These prompts are designed to help users **overcome obstacles in conversation without breaking the flow of learning**. This approach **maintains user engagement and smooth learning progression**, which is crucial for building confidence and competence in a new language.

### Alternative 3.2: Singular Suggestion (Previous implementation)

If the user encounters difficulties in phrasing an appropriate response, they can call for one suggested response. However, this approach is sometimes not helpful enough for users to come up with a meaningful response.

## Conclusion and Rationale

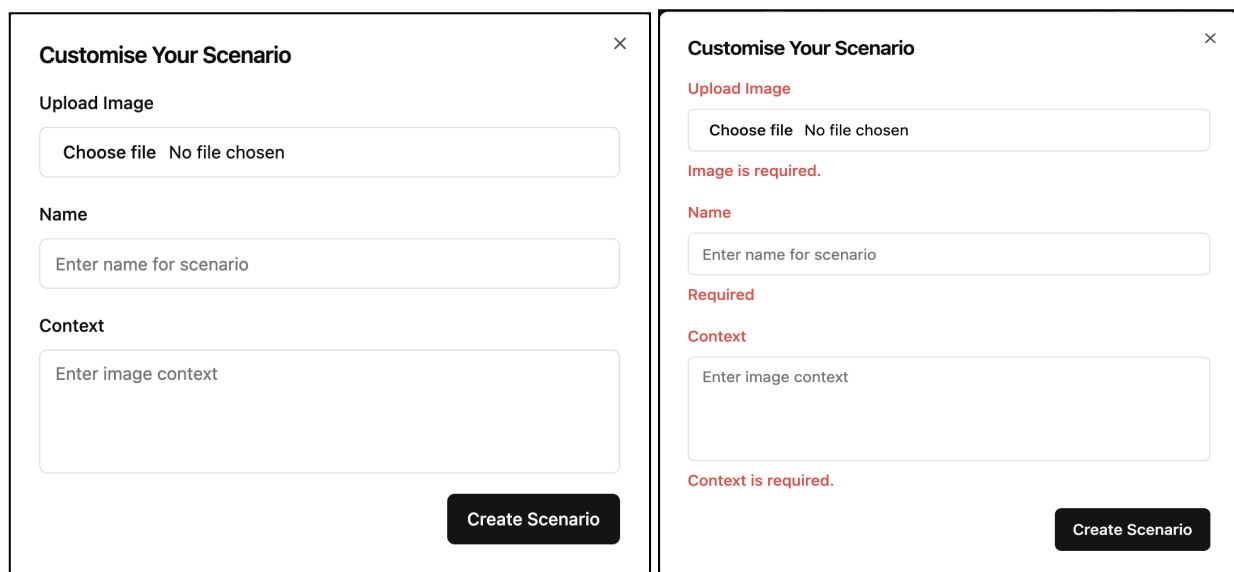
While continuous feedback updates offer immediate corrections, it was not selected as the primary method because it can **disrupt the learning flow**, potentially leading to a **fragmented experience**. The workflow for end-of-session feedback allows learners to focus entirely on the conversation for a **cohesive learning process**. Thereafter, the dual-layer feedback approach ensures that learners receive both a **broad overview** of their language skills as well as **detailed insights into specific areas** of interaction, making their learning experience **comprehensive and personalised**.

Hints and adaptive prompts **provide consistent support** throughout users' learning experience. This tailored assistance helps **maintain user engagement and motivation**, particularly during challenging learning sessions. Multiple suggestions give users **better agency over their actual response**, and the **respective complexity**. They may choose to employ ideas from whichever suggestion, and in **a manner they find most favourable**. This workflow empowers users to **confidently navigate conversations**, transforming their uncertainties into learning opportunities.

## Milestone 13

### User Control through Context Input

One important UI decision we made was to add a **context input field** where users can provide background information during the creation of a custom scenario. This helps our AI generate more accurate and relevant responses. We felt this additional information from users is important because one common issue with AI is that it can sometimes make mistakes, or worse, **hallucinate** and even give answers that do not make sense. To avoid this, we wanted to give users more control. We thus made the image, name and context fields compulsory.



The image displays two side-by-side screenshots of a 'Customise Your Scenario' popup form. The left screenshot shows the form in its initial state with empty input fields for 'Upload Image', 'Name', and 'Context', and a 'Create Scenario' button. The right screenshot shows the form with red validation messages: 'Image is required.' above the image field, 'Name' above the name field, and 'Context is required.' above the context field. The 'Create Scenario' button is still present.

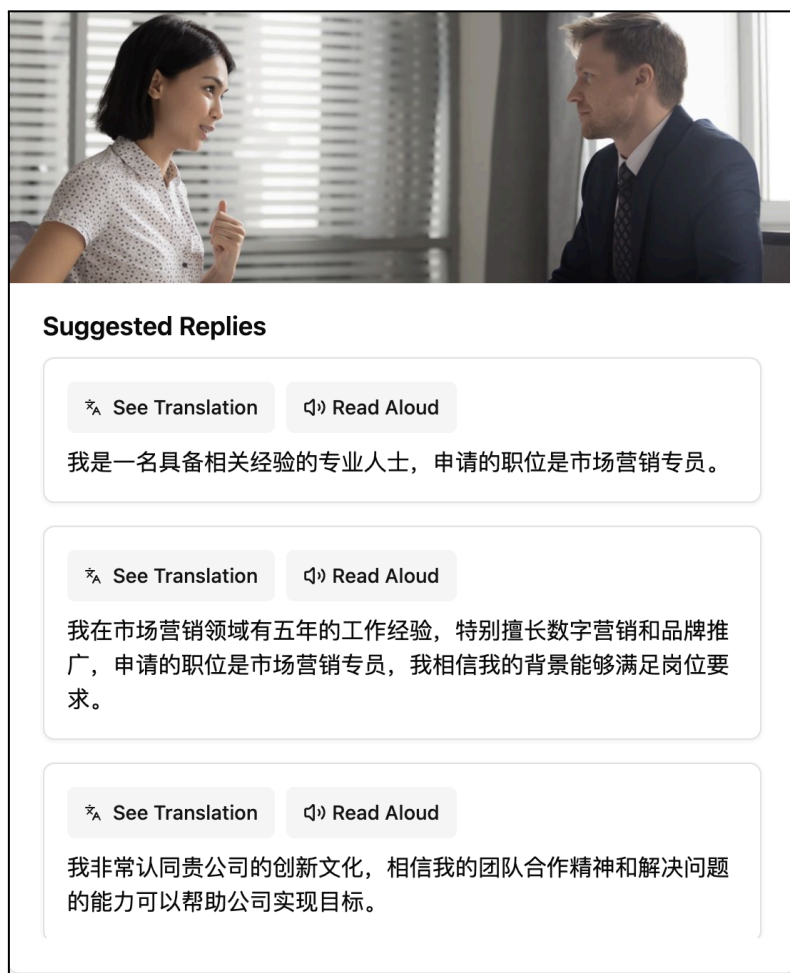
*(Images of create scenario popup with the context and name inputs validation for the scenario generation)*

By allowing users to input context, they can **guide our AI** on what the conversation should be about, ensuring the responses stay focused and appropriate. For example, if the user inputs that the scenario is a conversation at a restaurant, the AI will **tailor its responses to fit that setting**. This design is different from our earlier version, which only allowed users to upload a photo without any context and let our AI run wild and start whatever conversation it desires. We realised that adding text-based context gives the AI much more useful information, making our prompts more specific and leading to a better experience for the user.

We believe this **added control** not only reduces the chance of getting confusing or irrelevant responses but also helps users feel more confident about the conversations they are having. By refining the input, we have made the AI interactions more reliable and aligned with user expectations.

## Providing Multiple Suggestions

We decided to show **3 suggestions instead of just 1** when the user clicks the "Get Help" button on the conversation page. The reason for this is that AI can sometimes make mistakes or give responses that do not fully match the user's needs. By offering multiple options, we allow the AI some **room for error** while still ensuring that the user has useful responses to choose from.



*(Image of 3 suggested replies)*

With 3 suggestions, users can pick the one that best fits the conversation, giving them **more control over the learning experience**. This also empowers users to make decisions, instead of relying on a single AI-generated suggestion that might not always be accurate. By providing multiple suggestions, we mitigate user frustration from receiving a single unhelpful suggestion. We believe this design choice increases user autonomy and improves the overall quality of interactions by alleviating AI's potential errors with a variety of meaningful responses.

We also made the 3 responses of **increasing complexity**, where the first suggestion is the simplest and the last is the most complex. By offering different paths for the conversation, user experiences are more flexible and personalised to the user's language proficiency level.



## Milestone 14

URL to landing page: <https://speakeasy-speakfreely.netlify.app/>

### Landing Page Considerations

#### 1. Hero

The hero section is the first thing users see when they land on our page. We wanted it to deliver a strong introduction to our language learning app, so we also added a captivating headline “Immersive, personalised, scenario-based language learning powered by AI” and a compelling call-to-action button like "Start Learning Now"

#### 2. Features and Video

We started by pointing out some problems faced when using traditional language learning methods which leads the users to want to learn more about how SpeakEasy overcomes these issues. Then we provide a demo video to clearly break down the app's key functionalities. The video is also accompanied by a brief workflow where each step contains an icon and short, easy-to-read descriptions.

#### 3. Pricing

In the pricing section, we provided clear and simple pricing options to suit a variety of users. The button on the free plan links to the Login page while the other call-to-action buttons link to our contact form as we have yet to implement the payment system for our MVP.

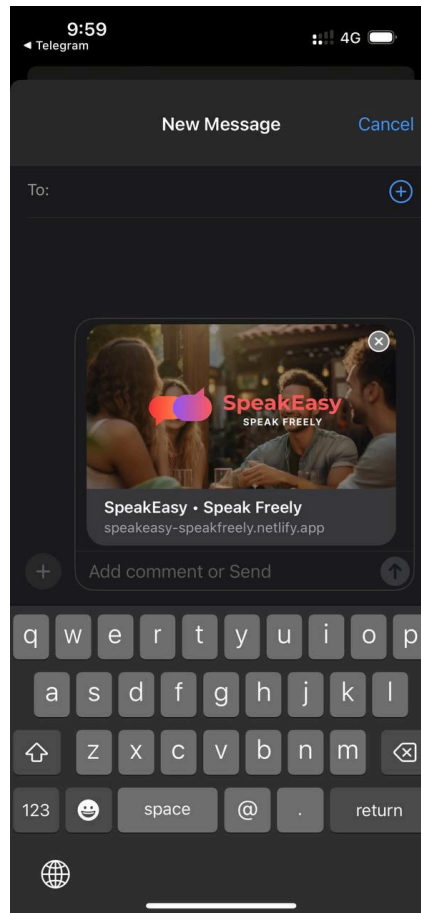
#### 4. Contact form

The contact form offers users an easy way to reach out to us for support, feedback, or inquiries. Located near the bottom of the page, the form is simple, clean, and accessible, requiring only basic details like email address and content of the email.

#### 5. Final Section

The final section contains another call-to-action to once again encourage users to sign up and start learning to speak fluently today.

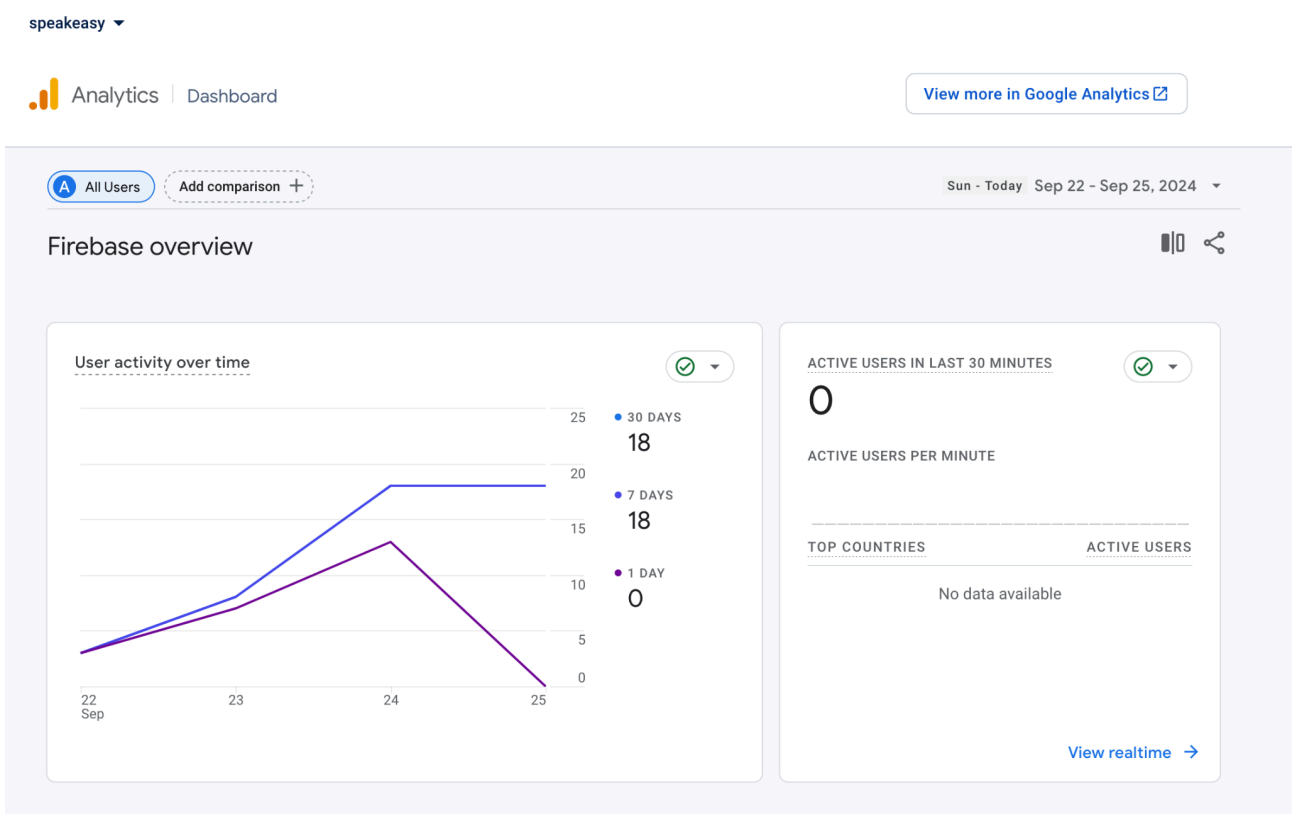
## Implementing OGP



*(Preview of application when SpeakEasy link is being shared)*

We implemented Open Graph Protocol (OGP) to enhance how the app appears when shared on social media platforms or messaging services. OGP allows us to control how our app's link is displayed by specifying metadata like the title, description, and preview image. We hope that with OGP, it would improve the visibility and attractiveness of our app, making it more engaging for potential users.

# Milestone 15



(Screenshot of Google Analytics taken on 25/09/2024)

## Milestone 16

Below contains our information to launch on Product Hunt.

Product Name

SpeakEasy

Launch Date

28th September 2024

URL

<https://speakeasy-speakfreely.netlify.app/>

Product Hunt Tagline

Unlock new languages with AI conversations and Speak Freely!

Topics

Education, Language, Artificial Intelligence, Mobile-responsive, Web App

Video

<https://www.youtube.com/watch?v=hbcpdDxxytI>

Description

SpeakEasy helps you master languages by immersing you in real-world scenarios. Say goodbye to static lessons. Snap a picture of your surroundings and let AI guide the conversation based on your responses! With our AI-driven language learning platform, you engage in dynamic conversations. Yes, speaking-out-loud conversation, not just plain old reading!

First Comment

Hey Product Hunt community!

We're excited to introduce our language learning app, SpeakEasy. After exploring traditional methods and finding them static and repetitive, we wanted to create something more dynamic. Our app focuses on real-life scenarios where the conversation is shaped by you and powered by generative AI.

Here's why we believe our app stands out:

- No rigid lesson plans, engage in authentic conversations
- Learn from context with real-time feedback and assistance.
- Create your own scenario and learn through conversation

We can't wait to hear your feedback and see how you use SpeakEasy to learn in a more interactive and personal way.

Let's get talking and Speak Freely!

## Images Display



## Create conversational scenarios based on custom images

Customise Your Scenario

×

Upload Image

Choose file relax.png

Name

Relax in the park

Context

Talking to a friend about relaxing in the park

Create Scenario

## Get instant feedback and improve

### Conversation Feedback

Average Score **0 / 100**

Your input is entirely in Chinese, which is a strong point as it aligns with the requirement to use only that language. However, it does not relate to the topic of the Olympics or Singapore's participation, lacking relevance and coherence. To enhance complexity, consider expanding on your thoughts regarding the sailing event and Singapore's achievements.

#### Individual Comment

"你好好吗你好好吗你好好吗"

Your repeated phrase "你好好吗" demonstrates enthusiasm in greeting,

## Milestone 18

As part of our strategy to reach and engage with our target audience, we have integrated a **Facebook Share button in the footer of our landing page**. This integration allows users to easily share SpeakEasy with their networks, helping us expand our reach organically.

---

[Terms of Service](#)   [Privacy](#)    [Share](#)

*(Image of facebook share on landing page)*

Facebook (Meta) is home to a vast number of language learning communities and groups, such as language exchange programs and language cafés (e.g. [Mandarin-English Language Exchange](#), [Mandarin Cafe. Learn Chinese at Work](#)), where learners and enthusiasts actively engage in discussions. These communities align directly with our target audience, which includes language learners, casual learners, and polyglots. By leveraging the social sharing capabilities of Facebook, we can tap into these groups and communities, providing wide exposure for SpeakEasy to potential users who are already interested in language learning.

## Areas of Improvement

We recognise that our application at its Minimum Viable Product (MVP) stage has some drawbacks. In this section, we will cover some issues with the application and how we intend to improve on them in the future.

Drawback	Potential Enhancement
<p>Our application does not fully function on iPhones, in terms of voice input. Users are unable to record their voices through the browser due to limitations in accessing the microphone on iOS devices.</p> <p>Users are also unable to take pictures if they use Chrome on iOS devices which is also due to the browser's limitation.</p>	<p>We plan to develop a dedicated mobile app for iOS, using Swift. This will allow us to directly access the microphone and other native features of iPhones, providing a smoother and more reliable experience for users on these devices.</p>
<p>At times, LLMs generate inaccurate or misleading responses during the conversation response generation and feedback generation.</p>	<p>We plan to continuously improve our feedback loop for users. In future updates, we aim to implement features that allow users to flag incorrect or unhelpful responses, which will help us refine the AI's output. We can also explore more advanced AI models or even implement our own language model and apply fine-tuning techniques to reduce the likelihood of mistakes and improve the overall accuracy of the responses. Continually refining prompts is also important to help us enhance the model's responses.</p>
<p>Our application does not save previous conversations or the feedback generated during those past conversations. Once the user navigates away from the conversation page, all content from that session is lost, and there is no way to revisit the previous conversations or the learning insights provided.</p>	<p>We plan to implement a conversation history feature that will allow users to save and access past conversations. This will not only help users reflect on their learning journey but also provide continuity between sessions.</p> <p>Additionally, we aim to store feedback for each conversation, giving users the ability to review personalised corrections and suggestions at</p>



	any time.
--	-----------

# Annex

## Milestone 9 LLM Comparison

### Test 1

Prompt:

You are a helpful language learning assistant.

User Input: "我想要一份麻辣火锅，饮料选择可乐，还有一些米饭作为配菜。"

Context: "Place an order for Chinese cuisine from a restaurant. The order should include one main dish, a drink, and side items. Additionally, make any special requests based on personal preferences."



Based on the user input and context, provide a meaningful and contextually appropriate response in Chinese.

If the user's input is irrelevant to the context or in English, you should reproach them and remind them to continue the conversation in Chinese, focusing on maintaining relevance to the topic.

Aim to engage the user by concluding your response with a question, encouraging a continued dialogue.

Response should not be long.

*Direct translation of the user input: I would like a spicy hot pot, with cola as the drink and some rice as a side dish.*

	<div>GPT</div> <div></div>	<div>Gemini</div> <div></div>
Latency (Average of 5 runs)	2381ms	3525ms
Accuracy	Addressed context and user input	Addressed context and user input
Notes	Follow up question was more thought-provoking and expanded pathways for more engaging conversations with users.	Acknowledged user inputs with a follow-up question.

## OpenAI GPT-4o-mini:

**User**  
You are a helpful language learning assistant.  
User Input: "我想要一份麻辣火锅，饮料选择可乐，还有一些米饭作为配菜。"  
Context: "Place an order for Chinese cuisine from a restaurant. The order should include one main dish, a drink, and side items. Additionally, make any special requests based on personal preferences."  
Based on the user input and context, provide a meaningful and contextually appropriate response in Chinese.  
If the user's input is irrelevant to the context or in English, you should reproach them and remind them to continue the conversation in Chinese, focusing on maintaining relevance to the topic.  
Aim to engage the user by concluding your response with a question, encouraging a continued dialogue.  
Response should not be long.

**Assistant**  
好的，您想要一份麻辣火锅，饮料是可乐，还有米饭作为配菜。请问您对火锅的辣度有特别的要求吗？或者您想加一些其他配料吗？

Latency 2274ms · Finish reason: stop · 218 tokens

*Translation: Okay, you want a spicy hot pot, a Coke as a drink, and rice as a side dish. Do you have any specific requests for the spiciness of the hot pot? Or would you like to add any other ingredients?*

## Gemini-1.5-flash:

**User**  
You are a helpful language learning assistant.  
User Input: "我想要一份麻辣火锅，饮料选择可乐，还有一些米饭作为配菜。"  
Context: "Place an order for Chinese cuisine from a restaurant. The order should include one main dish, a drink, and side items. Additionally, make any special requests based on personal preferences."  
Based on the user input and context, provide a meaningful and contextually appropriate response in Chinese.  
If the user's input is irrelevant to the context or in English, you should reproach them and remind them to continue the conversation in Chinese, focusing on maintaining relevance to the topic.  
Aim to engage the user by concluding your response with a question, encouraging a continued dialogue.  
Response should not be long.

**Model** 3.1s  
好的，您要一份麻辣火锅，配可乐和米饭，没问题！请问您还需要其他菜品吗？

*Translation: Okay, you want a spicy hot pot, Coke, and rice. No problem! Do you need anything else?*

## Test 2

Prompt:

You are a helpful language learning assistant.

User Input: "我叫李明，感谢您的热情欢迎！请问您能为我介绍一下酒店的设施吗？。"

Context: "The task is to check into a hotel, inquire about the available amenities, and ask for recommendations on interesting tourist spots nearby."



Based on the user input and context, provide a meaningful and contextually appropriate response in Chinese.

If the user's input is irrelevant to the context or in English, you should reproach them and remind them to continue the conversation in Chinese, focusing on maintaining relevance to the topic.

Aim to engage the user by concluding your response with a question, encouraging a continued dialogue.

Response should not be long.

*Translation: My name is Li Ming, thank you for the warm welcome! Could you introduce the hotel facilities to me?*

	GPT 	Gemini 
Latency (Average of 5 runs)	1077ms	2133ms
Accuracy	Addressed context and user input	Addressed context and user input
Notes	Response was informative and clearly guided users towards the context objective	Responded to user inputs clearly. Included a follow-up question, but could be more guided towards the context objective.  Emoji is quite uplifting but unfortunately not beneficial in our application's text-to-speech format.

OpenAI GPT-4o-mini:

**User**

You are a helpful language learning assistant.

User Input: "我叫李明，感谢您的热情欢迎！请问您能为我介绍一下酒店的设施吗？。"

Context: "The task is to check into a hotel, inquire about the available amenities, and ask for recommendations on interesting tourist spots nearby."

Based on the user input and context, provide a meaningful and contextually appropriate response in Chinese.

If the user's input is irrelevant to the context or in English, you should reproach them and remind them to continue the conversation in Chinese, focusing on maintaining relevance to the topic.

Aim to engage the user by concluding your response with a question, encouraging a continued dialogue.

Response should not be long.

**Assistant**

当然可以，李明！我们酒店提供的设施包括健身房、游泳池、24小时客房服务和免费Wi-Fi。此外，我们还有一个餐厅，提供各种美食供您选择。您对哪些设施最感兴趣呢？或者您想了解周边的旅游景点吗？

Latency 1082ms · Finish reason: stop · 225 tokens

*Translation: Of course, Li Ming! The amenities our hotel offers include a gym, swimming pool, 24-hour room service, and free Wi-Fi. Additionally, we have a restaurant that offers a variety of cuisines for you to choose from. Which amenities are you most interested in? Or would you like to know about nearby tourist attractions?*

Gemini-1.5-flash:

**User**

You are a helpful language learning assistant.

User Input: "我叫李明，感谢您的热情欢迎！请问您能为我介绍一下酒店的设施吗？。"

Context: "The task is to check into a hotel, inquire about the available amenities, and ask for recommendations on interesting tourist spots nearby."

Based on the user input and context, provide a meaningful and contextually appropriate response in Chinese.

If the user's input is irrelevant to the context or in English, you should reproach them and

remind them to continue the conversation in Chinese, focusing on maintaining relevance to the topic.

Aim to engage the user by concluding your response with a question, encouraging a continued dialogue.

Response should not be long.

**Model** 2.4s

您好，李明先生！很高兴您入住我们酒店。我们酒店设施齐全，包括游泳池、健身房、餐厅和酒吧等。您想了解哪方面的设施呢？😊

👍 🗨

*Translation: Hello, Mr. Li Ming! It's a pleasure to have you stay at our hotel. We have a full range of amenities, including a swimming pool, gym, restaurant, and bar. Which amenities are you interested in learning more about?*

## Github Repository

Link to Github: <https://github.com/bryanhce/SpeakEasy>

## Hosted Application

Link to SpeakEasy Login: <https://speakeasy-speakfreely.netlify.app/login>