INNOVATION IN LANGUAGE LEARNING International Conference

Designing an AI-Orchestrated Language Learning Platform with Adaptive Conversational Support

Michal Macinka[1], Stepan Hubalovsky[2]

[1,2]University of Hradec Kralove, Czech Republic

Abstract

Speaking anxiety is a well-known barrier in foreign language learning, often inhibiting students from practicing oral communication. This paper introduces a mobile application that enables learners to practice conversation with an AI partner in a low-pressure, self-paced environment. By offering immediate, judgment-free interaction, the AI conversational agent helps reduce the fear of speaking and lowers language anxiety. Recent studies have shown that technology-enhanced practice can alleviate learners' speaking anxiety and boost their confidence in using the language [1]. The approach is particularly beneficial for beginners in the early stages of language acquisition who may otherwise shy away from real-life conversation. The application's design is informed by the Common European Framework of Reference for Languages (CEFR) and its "Can-Do" statements, encouraging learners to engage in self-assessment and set personalized learning goals [2]. Aligning practice tasks with CEFR descriptors not only provides clear, attainable objectives but also fosters learner autonomy and motivation through goal-oriented progress tracking. As described by Kacetl and Klímová, mobile learning platforms can enhance students' autonomy and confidence by providing personalized practice opportunities [3], and our AI-driven app builds on these strengths. Similarly, a recent systematic review highlights that AI chatbots can create engaging conversational practice without social pressure, leading to improvements in speaking performance and learner engagement [4]. By integrating CEFR-based can-do goals with an adaptive AI conversation partner, the proposed solution supports self-directed learning and helps learners overcome the psychological barriers to speaking. This innovative approach demonstrates the potential of conversational AI to complement language instruction, providing learners with accessible, anxiety-reducing speaking practice that enhances their communicative competence.

Keywords: *Language Learning, Conversational AI, CEFR, Speaking Anxiety, Self-directed Learning*

1. Introduction

Speaking anxiety constitutes a pervasive hindrance to learning foreign languages [5]. Most learners tend to avoid oral communication, as they are significantly affected by the fear of making mistakes, negative evaluations, and not being exposed to enough real conversations. This anxiety is often evaluated with tools like the Foreign Language Classroom Anxiety Scale (FLCAS) and has been negatively linked with performance and motivation. Mobile learning and AI chatbots have opened up opportunities in efforts to alleviate these barriers. Research shows that conversational AI can provide a low-pressure and judgement-free environment in which learners can engage with practice [1], [4]. A systematic review of AI chatbots reveals that learners perceive these environments as friendly and free from anxiety, which in turn leads to increased confidence and readiness to present their ideas [6]. Similarly, a study of Google Assistant among adolescent EFL learners found that interaction over the course of 2 weeks significantly increased willingness to communicate and decreased speaking anxiety [7]. Another quasi-experimental study investigating the use of ChatGPT in writing practice found that students exhibited less anxiety due to the immediate feedback [8].

Despite these advancements, most applications developed so far employ a single conversational agent and do not utilize multiple role-plays or adapt to different target languages. The purpose of this resource is to design a platform that enables both English and Japanese learners to participate, coupling tasks with CEFR and RFJLE proficiency descriptors, and employing multi-agent orchestration to strike a balance between accuracy and fluency. The technical design of such a platform is discussed in this article rather than its experimental evaluation. Our architecture is motivated by insights from cognitive science, second language acquisition, and AI systems. Section 2 surveys related work in learning technologies and multi-agent frameworks. Section 3 provides a detailed description of the system architecture, covering the client-server infrastructure, the LangGraph orchestrator, and the roles of the agents. Section 4 reports on the adaptive dialogue flows in which

CEFR and RFJLE are operationalized, while Section 5 sketches an example use case with English and Japanese. Section 6 is summarized with the help of two figures: the first illustrates a flow of conversation with orchestrated agents, while the second describes the general architecture of the platform. Section 7 ends with a discussion and possible future research directions.

2. Related Work

2.1 Language Learning Anxiety and Conversational AI

Foreign language speaking anxiety typically arises when a learner fears embarrassment, unfavorable evaluation, or misunderstanding during the oral act of communication. Traditional classroom styles seem to exacerbate this anxiety further, as students' speaking time is minimal and subject to peer judgment. Mobile learning apps have been suggested to provide mobile practice opportunities. For example, Kacetl and Klímová [3] emphasize that smartphone-based language learning promotes learner autonomy at any time and in any place. Technology-enhanced language learning systems have also been shown to reduce public speaking anxiety in EFL settings [1]. According to Du and Daniel [4], AI-powered chatbots for English conversation can produce improvements in speaking performance and learner engagement. On the other hand, other comparisons suggest that AI-based speaking tasks may not always be more anxiety-reducing than peer-mediated ones, highlighting the need for careful design and maintenance of balance.

2.2 CEFR, RFJLE and Multilingual Frameworks

The CEFR is a common criterion that assesses language proficiency in listening, reading, speaking, writing, and interaction skills. Six levels (A1 to C2) are outlined, accompanied by Can Do statements that describe learners' abilities at each level. CEFR has been adapted for many languages, while the Referencing Framework for Japanese Language Education (RFJLE) offers the same guiding principles for learners of Japanese. These two systems were integrated into AI systems in a way that ensures a systematic pathway and objective assessment of the work done on learners. According to the Council of Europe, learning tasks should be aligned with the CEFR descriptors to foster self-assessment and goal setting [2]. The Japan Foundation's JF Standard and the newer RFJLE are aligned with communicative tasks for describing language proficiency [9]. Existing applications such as Duolingo integrate CEFR levels; however, they do not offer dynamic and conversational multi-agent feedback.

2.3 Multi-agent Orchestration and LLM Frameworks

These days, conversational agents are mainly implemented using contemporary large language models (LLMs). Frameworks such as LangChain and LangGraph provide abstractions for chaining LLM calls and orchestrating multi-step workflows. LangChain primarily aims to sequence the application of tools and agent behaviors; LangGraph expands this idea by modeling workflows as stateful graphs, where nodes represent agents or functions and edges represent transitions between them. LangGraph is particularly suitable for multi-agent systems, as it enables developers to design complex dialogues in which different agents can collaborate or compete with one another. We apply LangGraph to coordinate the agents responsible for conversation, grammar correction, and goal monitoring. This approach aligns with contemporary research advocating for choreographed and modular AI systems in education.

3. System Design

3.1 Client–server Infrastructure

The suggested platform features a client-server architecture that enables scalability and cross-platform accessibility. The client interacts with learners through a mobile application built on Flutter, making it usable across multiple platforms. Teachers and administrators utilize the web-based CMS technology stack, Directus, to manage course content, track student progress, and set goals for conversations. An incoming call, once authenticated, passes through rate limits and is then forwarded to internal services via the gateway API, which serves as the single point of entry for these services. Communication between the client and server is encrypted using HTTPS, while the CMS manages the storage of user data within a PostgreSQL database. The server is orchestrated using Kubernetes and is globally distributed across regions (Europe, Japan, and the USA) for low latency and data locality. Audio recordings and transcripts are saved in an object storage system (S3-compatible).

3.2 Multi-agent Orchestrator with LangGraph

At the core of the server lies the multi-agent orchestration engine, which is implemented using LangGraph. This engine orchestrates LLM calls and manages interactions between domain-specific agents regarding conversation flow (Figure 1). Major components include:

1. **Automatic Speech Recognition (ASR):** This module converts the learner's speech into text. We use Whisper or an alternative LLM-based ASR module. The generated transcripts are handed over to subsequent components.
2. **Natural Language Understanding (NLU):** This module processes the transcripts, extracting intents, entities, and semantic structures. It presents its findings to the DM and the grammar agent.
3. **Dialog Manager (DM):** The DM tracks the conversation state, selecting which agent should act next. The DM relies on LangGraph's stateful graph structure to direct traffic among the nodes representing the conversation agent, grammar agent, and goal agent. Transitions are determined by error rates, hesitations, and the learner's progression along the can-do graph.
4. **Response Generation:** Generates contextually relevant responses using an LLM (i.e., GPT 5 or fine-tuned model). Receives instructions (e.g., to consider the style, tone, and target language level) from the DM while generating responses.
5. **Text-to-Speech (TTS):** Converts the generated responses into human-like, audible speech. For Japanese, we utilize neural TTS models in synchronization with apt prosody, while for English, we leverage high-quality TTS engines, such as Coqui TTS.
6. **LangGraph Coordinator:** The underlying graph runtime coordinates message passing among the agents, conversation history management, and triggering of common tasks such as logging, scoring, and analytics.

The purposes of the agents are:

- **Conversation Agent (Partner)** – This is the online persona of a friendly interlocutor character wherein the main task is to sustain natural dialogue, encourage learners to speak, and respond appropriately. It does not interrupt during conversations for every little mistake; on the contrary, at the end of the conversation, it summarizes the errors made and offers suggestions. Such a design fosters fluency and lessens anxiety.
- **Grammar Agent (Coach)** – Actively monitors the output of the learner in real-time. Recognizing grammatical and phonological errors from NLU output, it proceeds to offer corrections on the spot. The DM decides when this process of interruption should occur; if errors are too numerous or too persistent, the coach offers a quick correction of the utterance and then encourages the learner to repeat it. This agent primarily ensures accuracy.
- **Goal Agent (Objectives)** – Keep track of the ongoing can-do descriptor and verify whether the learner has accomplished needed tasks or not (such as successfully ordering food or asking for directions). When the learner deviates, it brings the conversation back to the agreed-upon topic.
- **L1 Helper (optional)** – If triggered, intervention by this agent consists of explanations or translations in the learner's mother language. For example, a Japanese speaker learning English may ask a system to provide a grammar explanation in Japanese; the system responds with a translation model or a bilingual dictionary.

The LangGraph orchestrator models these agents as nodes of a directed graph. Each node can call upon a specific language model, such that a grammar agent uses one with error correction prompts. The edges encode the transition according to the DM decision. This graph is configurable, allowing teachers to establish conversation flows and set the threshold for decision-making.

3.3 Data Persistence and Analytics

Conversations are recorded for analytics and research. A time series database stores transcripts, error annotations, markers of hesitation, and agent interventions. Progress dashboards available to learners and teachers are populated using summary statistics, such as the average number of words per minute and error rates. Additionally, the system conducts CEFR level assessments based on the percentage of can-do tasks completed.

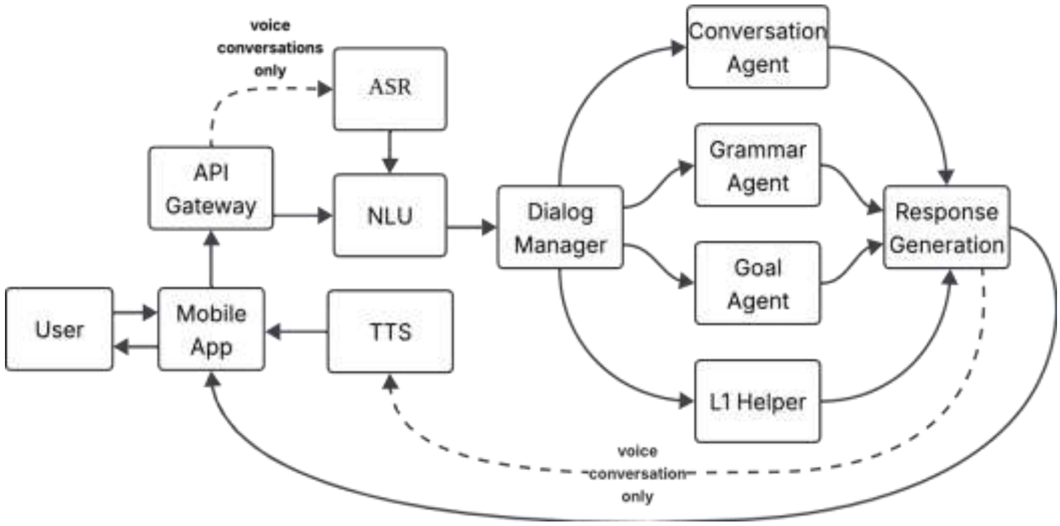

Fig. 1 Conversation flow. Through the mobile app, the learner's voice is captured by the API gateway, which relays it to the MCP server. ASR converts audio into text, while NLU parses the utterance. Depending on errors and progress, the dialogue manager routes through LangGraph, choosing either the Conversation agent, the Grammar agent, or the Goal agent. The selected agent invokes the relevant LLM, synthesizes the response via TTS, and delivers it back to the user. The L1 helper may be called upon whenever needed.

4. Aligning with CEFR and RFJLE

In ensuring teaching integrity, the platform links all conversations to CEFR and RFJLE descriptors. For example, CEFR descriptors imply learner capabilities at various levels in English. While A2 learners only learn to "ask and answer simple questions about familiar topics," B1 learners possess the ability to "deal with most situations likely to arise whilst travelling"; on Japanese, RFJLE provides parallel descriptors that capture cultural and pragmatic aspects. Our system uses these dimensions to structure conversation flows. The dialogue manager selects the task aligned with the current level of the learner and the student's previous performance. Where learners continually complete tasks with high accuracy and fluency, the system progresses to advanced tasks; otherwise, repeated mistakes provoke remedial tasks. The design ensures that learner progression happens while preventing errors from fossilisation. Teachers can create custom conversation graphs for specific curricula (e.g., business English or travel Japanese) via the CMS. Each node in the graph contains instructions for the agents, including target vocabulary, grammar points, and cultural notes.
Assessment is conducted in accordance with CEFR guidelines: the platform records whether the learner completes a task, utilizes target structures, and maintains coherence. It is sent to self-assessment dashboards and adds to formal standardised tests (e.g., TOEIC or JLPT). For Japanese learners, the system will also provide the JF-standardized Marugoto framework, which organizes tasks around specific themes, such as "shopping" or "hobbies."

5. Use Case: English and Japanese Learners

After registering, each learner completes a placement conversation with a bot. The bot asks simple questions (e.g., "Tell me about your daily routine") and uses CEFR/RFJLE descriptors to estimate the learner's level. Based on this, the Dialogue Manager (DM) selects appropriate tasks and agent modes:

- **Anna (Czech A2→B1, English):** Anna is assigned to a fluency practice. In this mode, the Partner agent engages in a natural conversation (e.g., ordering food at a café) and logs all errors (articles, prepositions, tense). It does not interrupt Anna; only at the end does it summarise her mistakes and suggest improvements.
- **Taro (Japanese B1→B2, English):** Taro's session alternates between fluency tasks and accuracy-focused tasks. When the DM detects repeated misuse of polite forms ("I want …" instead of "Could I …?"), it switches to the Coach agent, which intervenes after several errors,

explains the correct polite request, and has Taro repeat it. Once the correction is made, control returns to the Partner to continue the dialogue.

- **Miyu (Japanese B1→B2, English):** During a "booking a hotel room" task, Miyu repeatedly fails to grasp the concept of a "double room." The DM recognises the conversation is stalling and automatically triggers the L1 Helper. This agent translates the phrase into Japanese and offers a brief cultural note; Miyu does not have to activate it manually. The conversation resumes in Partner mode.

The system provides an overall summary at the end of each interval, covering task completion numbers, average words per minute, error types and frequencies, and personalized grammar exercises. Learners work through CEFR/RFJLE levels by completing various fluency and accuracy tasks.

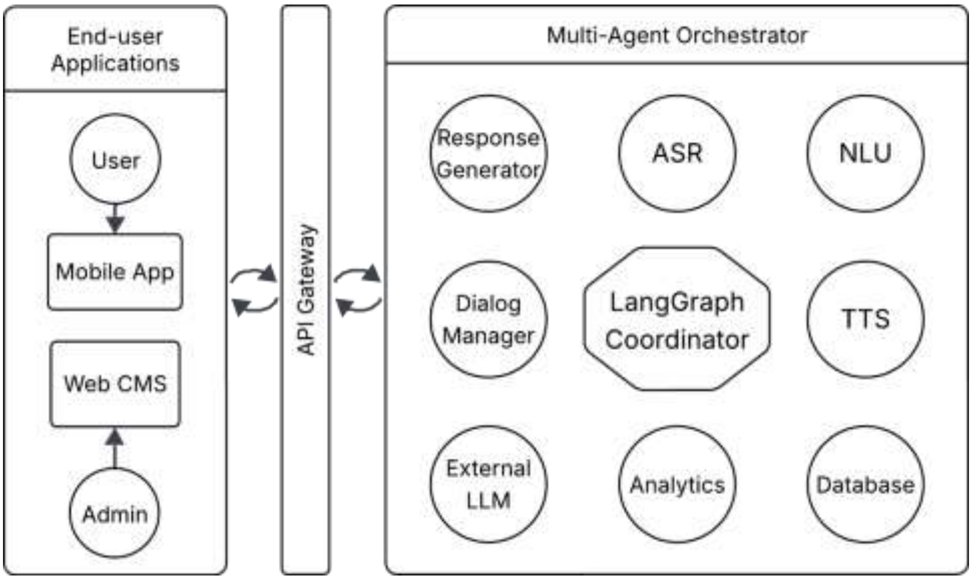

Fig. 2. Big picture architecture. Learners interact via the Flutter mobile app, which communicates with the API Gateway. Administrators manage content via the CMS. The gateway routes requests to the Multi-Agent Orchestrator, built with LangGraph, which contains modules for ASR, NLU, Dialogue Management, Response Generation, TTS, and the LangGraph Coordinator. The orchestrator calls external LLM and TTS services as needed. Data are stored in a database and analytics service for progress tracking.

6. Discussion

6.1 Academic and Technical Contributions

This study contributes further to the literature on AI-assisted language learning by introducing a multi-agent orchestration framework that strikes a balance between fluency and accuracy. The existing chatbots generally work with a single agent; our design utilizes LangGraph to orchestrate multiple specialized agents, enabling instant corrections and post-hoc reflections within the same session. The system also incorporates CEFR and RFJLE descriptors into the conversation flow to maintain its pedagogical perspective and facilitate cross-linguistic transferability. The architecture is scalable and modular, allowing new agents (e.g., pronunciation tutors or cultural coaches) to be easily added as nodes in the graph. From a technical viewpoint, using LangGraph provides strong state management and concurrency control, something that is less favorable in linear pipelines of LangChain.

6.2 Practical Implications for English and Japanese Learning

The adaptation to English and Japanese serves to illustrate the system's flexible nature. Japanese presents impediments, such as honorific speech (keigo) and topic-comment structure, which require the activation of sensitive feedback mechanisms. By integrating conversation and grammar agents, the system can instruct users on the proper use of polite forms and correct particle usage. The system, designed for English learners, will address more typically problematic areas, such as the use of articles, tense consistency, and the pronunciation of th sounds. The L1 Helper may provide explanations in Czech for English learners or in Japanese for Japanese learners to lessen their cognitive load. CEFR-aligned and RFJLE tasks ensure a systematic progression for learners and foster comparisons with standardized tests.

6.3 Limitations and Future Research

Several challenges remain pending. First, speech recognition errors are possible for non-native speakers, especially when speaking English with a Japanese accent or vice versa, which may later lead to incorrect feedback. Continued fine-tuning of ASR models will be necessary whenever needed for speech with accents. Secondly, a balance is critical in timing when the Grammar agent interrupts the conversation: too frequent corrections can discourage a learner, too few can allow errors to fossilize. This juggling act of concentrating on fluency versus accuracy has been receiving attention in recent work on the trade-offs of speech-driven language learning.
On the other hand, differential thresholding oriented towards learner profiles might alleviate that predicament. Thirdly, privacy and data protection will be implemented through anonymization and adherence to GDPR, as our platform currently relies on external LLM services. Fourthly, besides CEFR and RFJLE descriptors, accounting for different cultures in communication styles (such as Japanese indirectness) necessitates another consideration. Future studies will build upon the pilot studies conducted with English and Japanese learners, including a quantitative evaluation of anxiety reduction and learning gains, as well as an expansion to other languages.

7. Conclusion

This paper presents a design for an AI-orchestrated language learning platform that combines multiple agent roles through LangGraph to provide adaptive conversational practice. By aligning interactions with CEFR and RFJLE descriptors and offering optional support for mother tongue, the platform addresses both linguistic and psychological barriers. The architecture supports multilingual deployment and places a strong emphasis on scalability and privacy, thereby creating an environment for rigorous empirical evaluations. We expect that such a system will help learners overcome speaking anxiety, build confidence, and become more balanced in their overall language abilities. Future work will focus on further developing the adaptive mechanisms, validating the approach through larger user studies, and exploring integration with formal education settings.

ACKNOWLEDGEMENTS

The authors are grateful to the project of the Specific Research Faculty of Science UHK 2110/2025 for the financial support.

REFERENCES

[1] Chen Y. Effects of technology enhanced language learning on reducing EFL learners' public speaking anxiety. Computer Assisted Language Learning 2022;37:789–813. https://doi.org/10.1080/09588221.2022.2055083.
[2] Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Council of Europe Publishing; 2020.
[3] Kacetl J, Klímová B. Use of Smartphone Applications in English Language Learning—A Challenge for Foreign Language Education. Education Sciences 2019;9:179. https://doi.org/10.3390/educsci9030179.
[4] Du J, Daniel BK. Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. Computers and Education: Artificial Intelligence 2024;6:100230. https://doi.org/10.1016/j.caeai.2024.100230.
[5] Horwitz EK, Horwitz MB, Cope J. Foreign Language Classroom Anxiety. The Modern Language Journal 1986;70:125–32. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x.
[6] Li Y, Zhou X, Yin H, et al. Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review. Smart Learn Environ 2025;12:24. https://doi.org/10.1186/s40561-025-00379-0.
[7] Tai T-Y, Chen HH-J. The impact of Google Assistant on adolescent EFL learners' willingness to communicate. Interactive Learning Environments 2023;31:1485–502. https://doi.org/10.1080/10494820.2020.1841801.

[8] Hawanti S, Zubaydulloevna KM. AI chatbot-based learning: alleviating students' anxiety in english writing classroom. BullSocinfTheApp 2023;7:182–92. https://doi.org/10.31763/businta.v7i2.659.
[9] Japan Foundation. JF Standard for Japanese Language Education. Japan Foundation; 2010.