Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Chatbot Personalisation for EFL Learning: Integrating BKT and Task Context

Steve WOOLLASTON*, **Brendan FLANAGAN, & Hiroaki OGATA**
Graduate School of Informatics, Kyoto University, Japan
*s.m.woollaston@gmail.com

Abstract: English proficiency is critical in today's globalised world, yet many EFL learners face challenges in traditional classrooms due to large class sizes, rigid pedagogy, and varying proficiency levels. While AI chatbots offer accessible, low-pressure language practice, their effectiveness is limited by a lack of personalisation. This paper proposes a novel framework for dynamically personalising EFL chatbot interactions by integrating Bayesian Knowledge Tracing with semantic task analysis. The system aggregates learner data across multiple task-specific chatbots (e.g., translation, writing) to model vocabulary and grammar mastery. By analysing task context, the framework generates recommendations tailored to each learner, ensuring both proficiency-appropriate and activity-relevant support. This work advances personalised language learning by bridging the gap between isolated chatbots and a unified, adaptive learning ecosystem.

Keywords: Chatbot, conversational agents, EFL, chatlogs, second language learning, LLM, TELL, personalisation, bayesian knowledge tracing, semantic similarity

1. Introduction

English proficiency is increasingly important for living and working in today's interconnected world (Morita, 2017). This is especially true in Japan, which plays an active role in international economic and global political affairs. However, Japanese learners of English as a Foreign Language (EFL) face significant challenges within traditional classroom settings, including large class sizes, a reliance on grammar-translation pedagogical approaches, diverse proficiency levels among learners, and varying levels of motivation (Herawati, 2023). These and other factors can make it difficult to provide the individualised support and practice necessary for effective language acquisition (Zhang & Huang, 2024).

Educational technology, particularly the use of chatbots, has emerged as a promising tool to supplement traditional EFL instruction. Chatbots provide accessible and real-time interactive language practice opportunities outside the classroom, allowing students to learn independently and at their own pace in a friendly and low pressure environment. Chatbots can provide unlimited conversational practice, writing support, feedback, and other language learning activities (Brinegar, 2023).

1.1 *Personalisation*

Despite their advantages, the effectiveness of chatbots can be limited by their general one-size-fits-all approach (Zhang & Huang, 2024). Learners vary significantly in their current knowledge and learning pace. Personalised learning is a strategy to tailor learning content and experiences to the unique profile of each learner (Zerkowska, 2024). The US Department of Education similarly defines it as "...instruction in which the pace of learning and the instructional approach are optimised for the needs of each learner. Learning objectives, instructional approaches, and instructional content (and its sequencing) may all vary based on learner needs." (King & South, 2017, p. 9). AI chatbots have struggled to provide effective personalisation as they have little or no knowledge of the learner's current proficiency, and thus fail to adapt effectively to the prior knowledge and specific queries of individual learners. For EFL chatbots, personalisation means adapting interactions to the learner's Zone of Proximal Development (ZPD), providing appropriately challenging input (vocabulary and grammar), and ensuring relevance to the specific demands of the current task (Chaiklin, 2003). Such personalisation is widely endorsed for boosting learner engagement and attainment (Prain et al., 2013). Chatbot systems that can dynamically model learner progress and provide this targeted, individualised support is crucial for improving EFL practice and learning outcomes.

1.2 *Proposed Framework*

In this paper, we propose a novel personalisation framework designed to inform the responses of a suite of EFL learning chatbots currently deployed in high schools and a university in Japan. Each chatbot is task-specific, specialising in key language skills (active reading, translation, diary writing, and academic writing). Log data generated from these chatbots will be analysed in a unified backend system. This proposed system will employ Bayesian Knowledge Tracing (BKT) to model learners' mastery of vocabulary and grammar based on their interactions across all chatbots. By analysing task contexts (e.g., reading text excerpts, diary prompts and entries, translation prompts) via word embeddings, the system will generate real-time recommendations tailored to each learner's ZPD. These recommendations will target gaps in knowledge while ensuring relevance to the learner's current activity, such as suggesting vocabulary synonyms during translation practice or highlighting grammatical errors in diary entries.

1.3 *Contribution*

This research expands on Takii et al.'s (2021) vocabulary quiz recommendation system by introducing three key advancements to enhance personalised chatbot language learning. First, the framework enables **cross-chatbot learner modelling**, which builds a comprehensive, dynamic learner profile by aggregating interaction data from a variety of language learning activities like reading, writing, and translation, rather than being limited to a single task. Second, it is **task-contextualised**, combining proficiency estimates from Bayesian Knowledge Tracing (BKT) with semantic analysis of task content to ensure that recommendations are not only appropriate for the learner's skill level but also relevant to the specific activity. Finally, the system is designed with a **Japan-centric EFL alignment**, integrating nationally recognised CEFR-J grammar standards (Ishii & Tono, 2016) that are used in common school textbooks. This integration ensures both scalability and curricular relevance within the Japanese educational context.

2. Related Work

2.1 *Personalisation in Educational Technology*

Personalisation is a key goal in computer-assisted education, enabling educators to tailor instruction to meet the needs of diverse learners (Lin & Chang, 2023). This approach involves providing tailored recommendations and content based on areas where a learner struggles, which can be identified through their interactions with a system (Brinegar, 2023). Tailoring responses to learners' previous interactions can personalise learning experiences and potentially impact achievement (Huang et al., 2022). Examples include providing recommended reading materials based on learner characteristics like gender and interests, or tailoring explanations to a learner's proficiency level (Zhang & Huang, 2024). However, there are challenges in implementing effective personalisation. Despite positive effects reported in research, it can be unclear whether these effects result solely from the chatbot's personalised guidance or from the learners' strategic decisions (Lin & Chang, 2023).

2.2 *AI Chatbots in Language Learning*

Chatbots have gained prominence in language learning due to their ability to converse using natural language. They offer learners opportunities for dialogue-based practice and real-time feedback (Wang et al., 2024). While rule-based chatbots often struggled to understand nuanced input and produce natural language that was not formulaic or lacked contextual understanding, modern chatbots utilising large language models (LLM) are much more adept at producing natural varied responses and have a broad general knowledge (Woollaston, Flanagan, & Ogata, 2024). They are also robust to input errors, which is important in EFL contexts where mistakes and errors are common (Mzury, 2023). Chatbots provide numerous benefits in language learning:

- **Accessibility:** Learners can practise their language skills with chatbots anytime, anywhere, which is not easily possible with a human partner proficient in the target language. They offer quick access to support, addressing questions in real-time when human teachers are unavailable (Huang et al., 2022).
- **Lessened anxiety:** Chatbots offer a non-judgemental learning environment, are patient with errors and questions, and can provide a safe environment to experiment with new language without fear of making mistakes (Zhai, 2023).
- **Self-pacing and repetition:** Chatbots allow for self-paced conversation and learning (Ait Baha et al., 2023). They are also endlessly patient, providing unlimited practice - an important part of language learning (Huang et al., 2022).
- **Task-specific:** Chatbots can support various language learning activities, including conversation practice, answering language learning questions, conducting assessment and providing feedback, scaffolding writing skills, reading comprehension support, and translation practice.

2.3 *Learner Modelling*

Learner Modelling (LM) is the cornerstone of adaptive educational systems (ALS) and a critical component of personalised learning (Abyaa et al., 2019). Effective personalisation requires dynamic adaptation to a learner’s current knowledge and task-specific needs. This adaptation hinges on robust LM, which serves as the system’s "beliefs" about a learner’s knowledge, misconceptions, and progress (Kay et al., 2022). By continuously collecting and interpreting data—such as mastery and knowledge gaps—LM provides the foundation for tailoring instruction, feedback, and content sequencing to individual learners.

LM has focused on modelling knowledge and cognitive skills, leveraging educational theories like the Learning Curve and Forgetting Curve to track proficiency over time (Chen et al., 2017). Modern systems increasingly incorporate Open Learner Models, which make this data and insights to learners and teachers, fostering metacognitive awareness and self-regulated learning (Abyaa et al., 2019). More recently, the Open Knowledge and Learner Model (OKLM) has been proposed as a universal learner model framework that integrates Learning Analytics (LA) data from everyday learning activities with knowledge maps extracted from learning materials (Takii et al., 2024a). OKLM aims to enhance versatility and accuracy by linking learning activity logs to a knowledge map, allowing management and tracking of knowledge acquisition across diverse materials and contexts (Takii et al., 2024b).

2.4 *Relevance to Learning Context*

Cognitive psychology research shows that learning improves significantly when the content is directly relevant to the task (Bransford et al., 2000). This relevance boosts learner motivation and engagement by highlighting the practical value of the material (Ryan & Deci, 2000). It also fosters more meaningful and memorable learning as new information is integrated with existing knowledge within the specific context, ultimately improving retention (Jones, 1999). Furthermore, task-relevant learning content promotes more efficient learning by focusing on essential information. In EFL contexts, this principle underscores the importance of providing vocabulary and grammar recommendations that are directly applicable to the specific activity and context, thereby optimising the learning experience. Lexical and grammatical skills are best developed when relevant to the context, allowing for new information to be linked to prior knowledge (Jones, 1999).

By utilising semantic embeddings and distance metrics, language learning can be improved through the incorporation of meaning and context (Zerkowska, 2024). Semantic similarity measures, using models like Word2Vec, quantify the relatedness of text beyond keywords (Edapurath, 2022). These embeddings capture complex language nuances, enabling metrics like cosine similarity to determine contextual relevance (Hariprasath et al., 2024). This allows chatbots and personalised learning systems to provide accurate, meaningful responses and materials, going beyond simple keyword matching to infer context to a nuanced understanding.

3. Proposed Framework

3.1 *Chatbot Suite*

Several LLM-powered EFL chatbots have been developed and are currently deployed. **TAMMY** (Translation Assistant for MMasterY) is a chatbot designed to support English learning through translation tasks - a common task in English proficiency examinations in Asia (Ross, 2008). TAMMY allows learners to translate sentences between Japanese and English, receive feedback, ask questions, and engage in dialogue to improve understanding. In a pilot study, Woollaston et al., (2024) analysed the chat logs and a usability questionnaire. The findings indicated that TAMMY demonstrated good response validity and moderate success in guiding learners to accurate translations. Learners perceived the chatbot as friendly and easy to use, though their views on its usefulness and their intention to continue using it were neutral.

ARCHIE (Active Reading CHatbot for Interactive English) is a chatbot that scaffolds active reading strategies (e.g., summarisation, questioning, making connections) through 12 interactive activities (e.g., cloze passages, comprehension questions, role-play). Activities are tied to predefined texts (uploaded by the teacher), encouraging learners to engage deeply with content. Logs capture reading actions (e.g., page navigation, highlighting, memo-writing) and chatbot interactions. In a study to compare how high-proficiency and low-proficiency EFL learners interacted with ARCHIE, Woollaston et al., (2025) coded learner messages and employed Epistemic Network Analysis (ENA) to analyse learner behaviour with the assigned reading and the chatbot, They found that while all learners perceived the chatbot positively, high-proficiency learners interacted more frequently with the chatbot, exhibited more active reading strategies such as backtracking, and demonstrated less help-seeking behavior.

Penny and **Revision Rex** are writing support chatbots developed recently, and currently deployed at a Japanese high school and university, respectively. Penny assists with English diary writing. Each day, learners receive a prompt such as *"What is your favorite memory?"* or *"If you had one superpower, what would it be?"* After responding, they receive feedback and suggestions to improve their writing. Revision Rex supports academic writing in English, including essays and reports. It provides feedback and scaffolding in four key areas: vocabulary and word choice, surface features (spelling, punctuation etc.), grammar and sentence structure, and text organisation and cohesion.

Data about learner writing and interactions with each chatbot is stored in a database. With consent, these are stored securely and may be purged at any time upon request.

3.2 *System Architecture*

Figure 1 provides a summary of the proposed system architecture. Learner interactions and generated English text across different task-specific chatbots feed into a unified database. The system will employ Bayesian Knowledge Tracing (BKT) to model individual learner's vocabulary and grammar knowledge, drawing from log data and leveraging vocabulary and grammar preprocessing. The Recommendation Engine will then use this learner model, along with the task context, to tailor the chatbot's responses, aiming to provide personalised and relevant feedback and suggested next-learning steps aligned with the learner's ZPD.

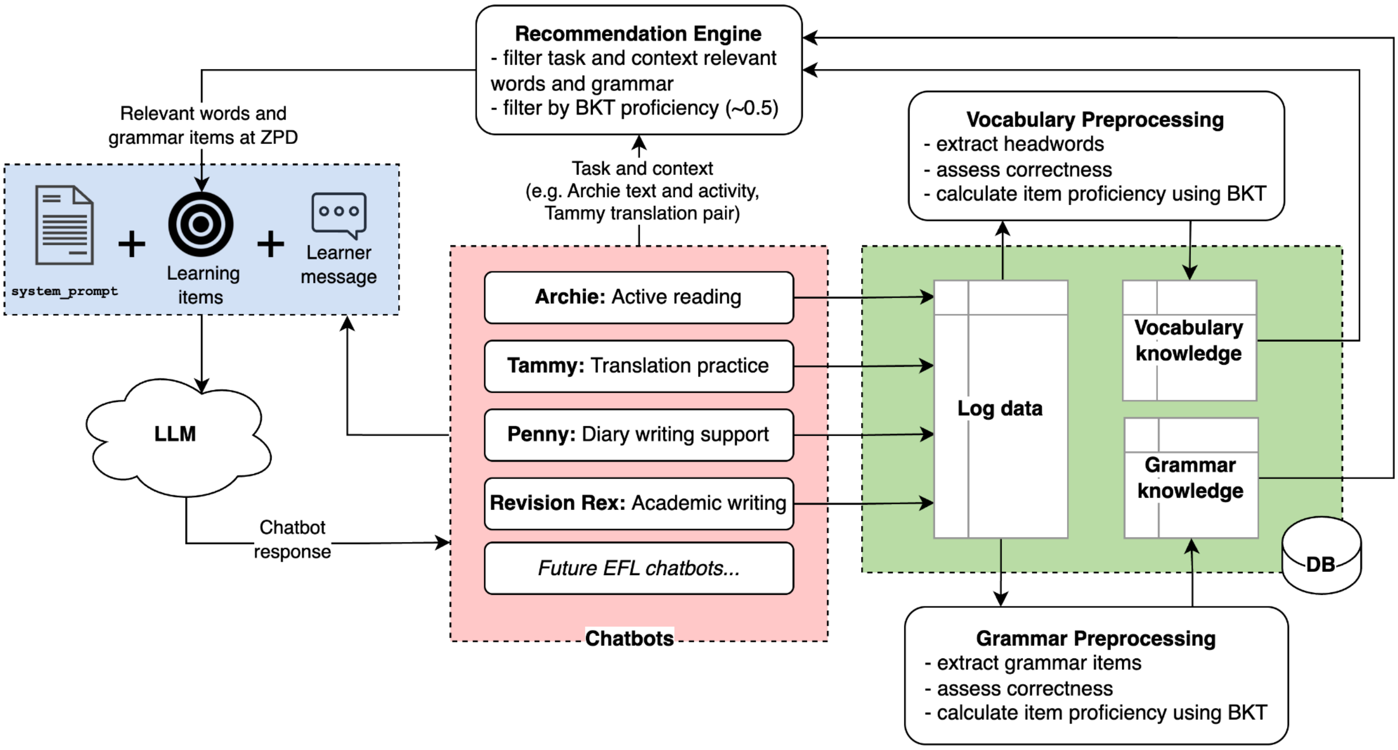

Figure 1. System architecture diagram

3.3 *Learner Modelling*

The system will dynamically model each learners' vocabulary and grammar knowledge by analysing learner-generated text from all chatbots. This includes messages sent to the chatbots and learner writing (e.g. diary entries in the diary chatbot Penny).

3.3.1 *Vocabulary Extraction and Evaluation*

For all learner submitted text, the system will execute a multi-step pipeline:
1. **Text Processing:** Inputs will be split into sentences, and each word will be checked for spelling errors. Misspelled words will be replaced with the most likely candidate (e.g., "acomodate" → "accommodate").
2. **Lexical Normalisation:** Each word will be lemmatised to its base form (headword) and tagged with its part-of-speech (POS). For example, "running" (verb) will be normalised to "run-VB," distinguishing it from noun or adjective uses.
3. **Usage Logging:** Each time a learner uses a normalised headword–POS pair, we record a log entry containing metadata such as the timestamp, the source chatbot, the activity context, and other relevant metadata (e.g., "ARCHIE; Summarisation; 'The Dangers of Social Media'"). We also save the learner's original sentence for context. As learners interact with the chatbots over time, each unique headword–POS pair accumulates one or more usage logs.

Whether a word is used correctly will be determined by evaluating two aspects:
- **Spelling:** Misspelled words will be automatically flagged as incorrect.
- **Contextual accuracy**: Correctly spelled words will be assessed for contextual accuracy using an LLM for classification. The model will evaluate whether the word aligns with the sentence's semantic and syntactic requirements (e.g., *Please <u>except</u> my invitation* vs. *Please <u>accept</u> my invitation*"). The model's classification (correct or incorrect usage) will be stored for later use by BKT.

3.3.2 *Grammar Extraction and Evaluation*

This section details the process of extracting and evaluating grammar usage in learner-generated text. This approach leverages the well-defined CEFR-J framework to identify specific grammatical structures and assesses the accuracy of their use within the context of the specific learning activity.

The Common European Framework of Reference for Languages (CEFR) provides a robust and widely accepted standard for describing language proficiency (Figueras, 2012). It provides a framework of language ability, from basic (A1) to fully proficient (C2), based on what language learners can do. In the Japanese context, the CEFR has been adapted into CEFR-J, which aims to apply the framework to the specific context of English language education in Japan, taking into account the unique challenges and needs of Japanese learners and aligning with Japanese educational policies and curriculum (Masashi, 2012).

A key component of CEFR-J is the identification of specific grammar items relevant to each proficiency level. These grammar items, derived from analyses of Japanese EFL textbooks and cross-referenced with other CEFR-aligned resources, are defined using regular expressions (regex). These regexes specify patterns based on word forms, lemmas, and parts of speech, enabling the automated extraction of grammar instances from English text. Table 1 provides a representative selection of these grammar items, illustrating their definition and corresponding example sentences. A comprehensive list of all 263 CEFR-J grammar items and their associated extraction regular expressions is publicly available at cefr-j.org (Ishii & Tono, 2016).

Table 1. *CEFR-J Grammar Item Examples*

CEFR-J Level	Grammar Item	Example sentence
A1.1	INDEFINITE ARTICLES (a/an + noun)	I have a book.
A1.1-A1.2	I am (... including question, negative)	I am happy. Am I next? I am not sad.
A1.1-B1.2	WH- QUESTION: How ADJ/ADV ...? (How + adjective/adverb ...?)	How quickly can you run?
A1.2-A1.3	TENSE/ASPECT: PRESENT PROGRESSIVE (present continuous)	She is reading.
A1.2-A2.2	VERB to DO (verb + to-infinitive)	I want to sleep.

Our grammar extraction and evaluation process involves the following steps:

1. **Sentence segmentation:** Learner-generated text is divided into sentences.
2. **LLM-enhanced error detection:** Each sentence is processed. Where grammar is problematic, an LLM will generate a grammatically corrected version with minimal changes, while preserving the original meaning. This step aids in highlighting potential errors for subsequent evaluation. Both the original and rewritten sentences are retained.
3. **Grammar item extraction:** For each sentence (or *minimally corrected* sentence), a list of all grammar items present in the sentence is generated using regex, as defined by the CEFR-J framework.
4. **Learner grammar tracking:** The system tracks learner grammar usage by logging each identified grammar item in their writing. For a learner's first use of a grammar point, it's recorded in their personal grammar record with its CEFR level. Subsequent uses trigger a usage log entry, noting the time, location (chatbot, activity, context), and an LLM-assessed correctness (correct/incorrect).

This method systematically identifies and evaluates grammar usage based on CEFR-J using LLM assistance for error highlighting and correctness assessment.

3.3.2.1. *English Vocabulary Profile*

While the CEFR-J framework was initially considered for vocabulary, its coverage (7,988 words, excluding C1–C2 levels) proves insufficient to accommodate advanced learners or support multilingual learners with diverse L1 backgrounds. To ensure inclusivity across proficiency levels and linguistic contexts, we will utilise the English Vocabulary Profile (EVP) as our predefined vocabulary list (Capel, 2015). The EVP offers two main advantages: complete coverage across all CEFR proficiency bands (A1-C2) and validation through extensive corpus analysis. This ensures our system can provide appropriate vocabulary recommendations for learners at any stage of their vocabulary learning.

3.4 *Modelling Vocabulary and Grammar Mastery*

To enable personalised and adaptive interactions across our suite of EFL chatbots, we will implement BKT to model learners' mastery of vocabulary and grammar in—or as close as possible to—real time. BKT's probabilistic framework will allow us to estimate the likelihood that a learner has mastered specific knowledge components (KCs)—such as a CEFR-J grammar rule or a vocabulary word—based on their observed performance during chatbot interactions. Unlike static pre-test assessments, BKT will dynamically update these estimates as learners engage with different tasks, ensuring that recommendations remain aligned with their evolving proficiency.

For vocabulary tracking, we will treat each lemmatised word–POS pair (e.g., "run-VB") as an individual KC. Since language knowledge is sensitive to forgetting over time, we will extend the classic BKT model to incorporate memory decay (BKT+F). This will allow us to adjust mastery estimates based on how frequently and recently a word has been encountered across different chatbot activities. Words that appear in multiple contexts will contribute to a more robust, generalised estimate of a learner's current knowledge.

The output of the BKT engine will be a continuously updated knowledge profile for each learner, which will drive personalisation across all chatbots. For vocabulary, this profile will prioritise words near the learner's ZPD—those with a $P(mastery)$ near 0.5—while ensuring relevance to their current activity (e.g., suggesting "leap" during a translation task if "jump" appears in their sentence). For grammar, the system will highlight errors on high-priority KCs, such as frequently missed A2-level rules during diary writing.

By modelling learner knowledge using BKT, our framework will advance beyond static rule-based personalisation, offering three key advantages:

1. Continuous updates to learner models as new data is generated across chatbots / tasks;
2. Effective management of errors, slips, and lucky guesses in learner language, and,
3. Scalability through shared KC definitions that unify data from diverse chatbots.

This approach will form the foundation for our system's ability to deliver tailored, context-aware support—a critical step toward addressing the personalisation gaps in current language learning chatbots.

3.5 *Aligning Task Relevance with Learner Model*

Our framework dynamically aligns learners' proficiency estimates (via BKT) with semantically relevant vocabulary and grammar suggestions through a three-stage pipeline:

1. **Proficiency Modelling:** BKT will track mastery probabilities for each knowledge component (KC)—lemmatised vocabulary (e.g., "run-VB") and CEFR-J grammar items (e.g., "past simple tense")
2. **Task Context Analysis:** When a learner engages in an activity (e.g., translating a sentence, drafting a diary entry), the system will extract keywords and grammatical structures from their input. Using word embeddings, the semantic similarity between these elements and the EVP will be computed. Grammar relevance will be ascertained by LLM. For example, in a translation task about sports, words like *compete* and *victory*

might be identified as contextually relevant vocabulary. For grammar, depending on the context (task and content), the system will suggest grammar items that are relevant (e.g. past tense usage during diary writing).

3. **ZPD Filtering:** The system will prioritise KCs with *P(mastery)* near 0.5 and semantic similarity >0.6 to task keywords, ensuring suggestions like "stroll" (B1) for "walk" (A2) during a writing task. When processing grammar, the mechanism will prioritise corrective feedback based on the relevance to the current activity, such as emphasising tense consistency in narrative writing or article usage in descriptive tasks. This methodology ensures recommendations are simultaneously tailored to the learner's developmental stage and the specific linguistic demands of their current activity.

3.6 *Integration into Chatbot Dialogue*

When the Recommendation Engine identifies relevant vocabulary or grammar items based on the learner's model and the task context, this information will be used to dynamically shape the chatbot's responses via prompting. For vocabulary, the chatbot might subtly introduce target words in revisions, as synonyms, or within explanations, encouraging exposure, comprehension, and usage in context. For grammar, the chatbot will provide corrective feedback on errors, offer rephrased sentences demonstrating correct usage, or pose targeted questions that guide the learner to apply specific grammatical structures. This integration within the natural flow of the chatbot interaction will ensure that recommendations are timely, contextualised, and directly relevant to the learner's immediate learning activity.

4. Discussion

4.1 *Conclusion*

The proposed framework addresses three key EFL chatbot challenges: fragmented data, contextual misalignment, and relevance and scalability to the Japanese EFL context. Traditional isolated chatbots fail to track holistic progress, so our system aggregates data across activities (translation, writing, active reading) to reflect more comprehensive language skill integration. We employ BKT for its dynamic, probabilistic mastery estimation, which incrementally adapts to evolving proficiency. Unlike generic feedback systems, our semantic analysis ensures task-relevant suggestions, leveraging evidence that contextual learning boosts retention. The CEFR-J and EVP standards were chosen to balance Japanese curriculum specificity with comprehensive vocabulary coverage (A1–C2). By prioritising integration with existing tools (e.g., TAMMY, ARCHIE), the design ensures practical adoption in the Japanese context.

4.2 *Limitations*

The proposed framework faces several challenges, the most significant of which is **privacy and data security**. Aggregated learner data requires stringent safeguards to protect their privacy, such as data anonymisation and access control. Another issue is the potential for **bias and hallucinations** in LLM-generated feedback, as it may inherit biases from training data or produce incorrect corrections, risking learner trust (Zhai, 2023). For example, a chatbot might misclassify a creative but valid use of grammar as an error. The chatbots' **narrow skill focus** presents another challenge, as they currently concentrate on productive literacy skills, primarily writing, while excluding critical communicative skills like speaking and listening. Future chatbots should aim to expand the range of skills they develop. Lastly, the **LLM dependency** for error detection introduces stochastic variability; their judgments may lack the consistency, nuance, and experience of human tutors, especially when dealing with ambiguous or culturally specific language.

4.3 *Future Work*

To further enhance the framework's efficacy and applicability, several development and research directions warrant exploration. First, the development of interactive dashboards would empower learners and teachers to monitor progress, adjust recommendation thresholds, and validate AI-generated feedback—addressing current transparency gaps and introducing explainability. Expanding into multimodal learning, such as speech recognition for speaking and listening exercises, will significantly expand the range of language skills that can be effectively modeled and supported. Additionally, mitigating biases in LLM outputs through systematic audits and human-AI review (human in the loop) may improve accuracy. Longitudinal studies in real-world classrooms are also critical to empirically validate the system's impact on proficiency gains compared to non-personalised approaches.

By iterating on this foundation system, we hope this framework will develop into a universal adapter for language learning tools, bridging the gap between isolated chatbots and a cohesive task-contextualised personalised language learning ecosystem.

Acknowledgements

This proposal was supported by Council for Science, 3rd SIP JPJ012347, and JSPS Grant-in-Aid for Scientific Research (B) JP20H01722 and JP23H01001, (Exploratory) JP21K19824, JP22KJ1914, (A) JP23H00505, and NEDO JPNP20006.

References

Abyaa, A., Khalidi Idrissi, M., & Bennani, S. (2019). Learner modelling: systematic review of the literature from the last 5 years. *Educational Technology Research and Development*, *67*(5), 1105–1143.

Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2023). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, *29*(8), 10153–10176.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded Edition, Vol. 11). National Academies Press.

Brinegar, M. (2023). Chatbots as a Supplementary Language Learning Tool: Advantages, Concerns, and Implementation. *International Journal of Education and Social Science Research*, *6*(6), 223–230.

Capel, A. (2015). The English vocabulary profile. *English Profile in Practice*, *5*(1), 9–27.

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. *Vygotsky's Educational Theory in Cultural Context*, 39–64.

Chen, Y., Liu, Q., Huang, Z., Wu, L., Chen, E., Wu, R., Su, Y., & Hu, G. (2017). Tracking knowledge proficiency of students with educational priors. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Conference on Information and Knowledge Management, Singapore.

Edapurath, V. N. (2022). Enhancing chatbot response relevance through semantic similarity measures. *Journal of Artificial Intelligence and Cloud Computing*, *1*(1), 1–5.

Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, *66*(4), 477–485.

Hariprasath, Paramarthalingam, A., Sundaramurthy, S., & Cirillo, S. (2024). A study on word embeddings in local LLM-based chatbot applications. *2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 232–237.

Herawati, A. (2023). *Personalized Learning in Teaching English as Foreign Language: Limiting the Challenges, Increasing Its Effectiveness*. 3–12.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot‐supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257.

Ishii, Y., & Tono, Y. (2016). A frequency survey of grammar items for the CEFR-J grammar profile. *Proceedings of the 22nd Annual Meeting of the Association for Natural*

Language Processing, 777–780.
Jones, C. (1999). Contextualise & personalise: Key strategies for vocabulary acquisition. *RECALL-HULL*, *11*(3), 34–40.
Kay, J., Bartimote, K., Kitto, K., Kummerfeld, B., Liu, D., & Reimann, P. (2022). Enhancing learning by Open Learner Model (OLM) driven data design. *Computers and Education: Artificial Intelligence*, *3*(2022), 100069.
King, J., & South, J. (2017). Reimagining the role of technology in higher education: A supplement to the national education technology plan. *US Department of Education, Office of Educational Technology*, *10*, 1–10.
Lin, M. P.-C., & Chang, D. (2023). CHAT-ACTS: A pedagogical framework for personalized chatbot to enhance active learning and self-regulated learning. *Computers and Education: Artificial Intelligence*, *5*, 100167.
Masashi, N. (2012). The development of the CEFR-J: Where we are, where we are going. *New Perspectives for Foreign Language Teaching in Higher Education: Exploring the Possibilities of Application of CEFR. Tokyo: Tokyo University of Foreign Studies*, 10–116.
Morita, L. (2017). Why Japan needs English. *Cogent Social Sciences*, *3*(1), 1399783.
Mzury, D. L. (2023). Exploring EFL learners' linguistic error and/or mistakes: Causes, and correction techniques. *Zanco Journal of Human Sciences*, *27*(4), 333–346.
Prain, V., Cox, P., Deed, C., Dorman, J., Edwards, D., Farrelly, C., Keeffe, M., Lovejoy, V., Mow, L., Sellings, P., Waldrip, B., & Yager, Z. (2013). Personalised learning: lessons to be learnt. *British Educational Research Journal*, *39*(4), 654–676.
Ross, S. J. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing*, *25*(1), 5–13.
Ryan, R. M., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78.
Takii, K., Flanagan, B., & Ogata, H. (2021). EFL vocabulary learning using a learning analytics-based E-book and recommender platform. *2021 International Conference on Advanced Learning Technologies (ICALT)*, 254–256.
Takii, K., Liang, C., & Ogata, H. (2024a). OKLM: Open Knowledge and Learner Model using educational big data. *International Conference on Computers in Education*, 711–714.
Takii, K., Liang, C., & Ogata, H. (2024b). Open knowledge and Learner Model: Mathematical representation and applications as learning support foundation in EFL. *International Conference on Computers in Education*, 605–614.
Wang, F., Cheung, A. C. K., Neitzel, A. J., & Chai, C. S. (2024). Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. *Review of Educational Research*. https://doi.org/10.3102/00346543241255621
Woollaston, S., Flanagan, B., Ocheja, P., Dai, Y., & Ogata, H. (2024). TAMMY: Supporting EFL Translation Practice With an LLM-Powered Chatbot. *Proceedings of the 32nd International Conference on Computers in Education*, *1*, 595–604.
Woollaston, S., Flanagan, B., Ocheja, P., Toyokawa, Y., & Ogata, H. (2025). ARCHIE: Exploring language learner behaviors in LLM chatbot-supported active reading log data with epistemic network analysis. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 642–654.
Woollaston, S., Flanagan, B., & Ogata, H. (2024). Chatbots and EFL learning: A systematic review. *Joint Proceedings of LAK 2024 Workshops*, 89–98.
Zerkowska, A. M. (2024). *Personalized Language Learning in the Age Of AI. Leveraging Large Language Models for Optimal Learning Outcomes* [University of Gothenberg]. https://gupea.ub.gu.se/handle/2077/81934
Zhai, C. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, *4*, 100134.
Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon*, *10*(3), e25370.