ATAIGI: An AI-Powered Multimodal Learning App Leveraging Generative Models for Low-Resource Taiwanese Hokkien

Yun-Hsin Chu[+] **Shuai Zhu**[†] **Shou-Yi Hung**[†] **Bo-Ting Lin**[+]
En-Shiun Annie Lee[†,‡,*] **Richard Tzong-Han Tsai**[+,§,*]
[+]Department of CSIE, National Central University, Taiwan [†]University of Toronto
[§]Center for GIS, RCHSS, Academia Sinica, Taiwan [‡]Ontario Tech University
cyh74562@gmail.com, sy.hung@mail.utoronto.ca

Abstract

Many endangered languages are at risk of extinction due to barriers in communication and generational gaps that hinder their preservation. A cause for languages becoming endangered is the lack of language educational tools and artificial intelligence (AI) models for these low-resource languages. To address this, we propose the ATAIGI learning app designed with AI-powered models leveraging multimodal generative techniques. Our app offers users a comprehensive learning experience by providing translated phrases and definitions, example sentences, illustrative images, romanized pronunciation, and audio speech to accelerate language learning. ATAIGI is built on five AI models that are rigorously benchmarked individually, with our Transliteration Model achieving state-of-the-art results for Taiwanese Hokkien transliteration. ATAIGI is available for all to learn the endangered language of Taiwanese Hokkien, an endangered language spoken in Taiwan. A human evaluation conducted demonstrates the effectiveness of ATAIGI in improving language proficiency and cultural understanding, supporting its potential for the preservation and education of endangered languages like the Taiwanese Hokkien. *App deployed using Netlify*[1]*, codebase available on GitHub*[2].

1 Introduction

Natural language processing tools and models used for language learning are commonly applied to high-resource languages (HRLs) rather than low-resource languages (LRLs) (Schwartz et al., 2004). Recent language learning tools showed they could be effective in helping users learn a language (Doi et al., 2021; Woo and Choi, 2021; Rebolledo Font de la Vall and Gonzalez Araya, 2023). Existing research (Kim et al., 2023) investigates the feasibility of *GPT* (Radford et al., 2018) as a learning tool for a second language; however even a model as powerful as *GPT* does not perform well for LRLs (Ghosh and Caliskan, 2023; Stap and Araabi, 2023; Robin et al., 2022). Despite the abundance of software tools and literature on HRLs available, LRLs are often overlooked due to the scarcity of data and models required to create learning tools.

Taiwanese Hokkien is an endangered LRL that represents a cultural heritage on the brink of extinction (Chen, 2023). A generational linguistic gap exists between the elderly who predominantly communicate in Taiwanese Hokkien, and the younger generation who communicate in HRLs such as Mandarin and English. Although useful tools such as the translation tool iTaigi[3] and the text-to-speech (TTS) tool TauPhahJi[4] are available, but the scattered and unintegrated nature of these resources makes it difficult for users to master the LRL. When using specific vocabulary, learners often face difficulties recalling suitable terms and examples quickly.

We created ATAIGI, which pronounces "learning Taiwanese Hokkien" in its native tongue, to serve as a learning framework for teaching the endangered language of Taiwanese Hokkien. Based on the mentioned shortcomings, the ATAIGI learning app makes the following contributions:

1) Developed a complete learning system in Taiwanese Hokkien (Figure 1) that provides translations of simple to understand phrases, and also examples of sentences, images, Romanized pronunciation, and audible speech to improve learning and memory retention.

2) Utilized five AI models to support a comprehensive Taiwanese Hokkien learning framework. Two models were specifically trained for our pro-

[*]Corresponding authors
[1]https://hokkien-translation.netlify.app/
[2]Our repository is under MIT license, available at https://github.com/HokkienTranslation/HokkienTranslation
[3]iTaigi is a dictionary introducing Taiwanese local language terms, documenting modern expressions used in Taiwan.
[4]TauPhahJi converts Taiwanese Hokkien Hanzi into its romanized form and then into speech.

Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations), pages 11–19

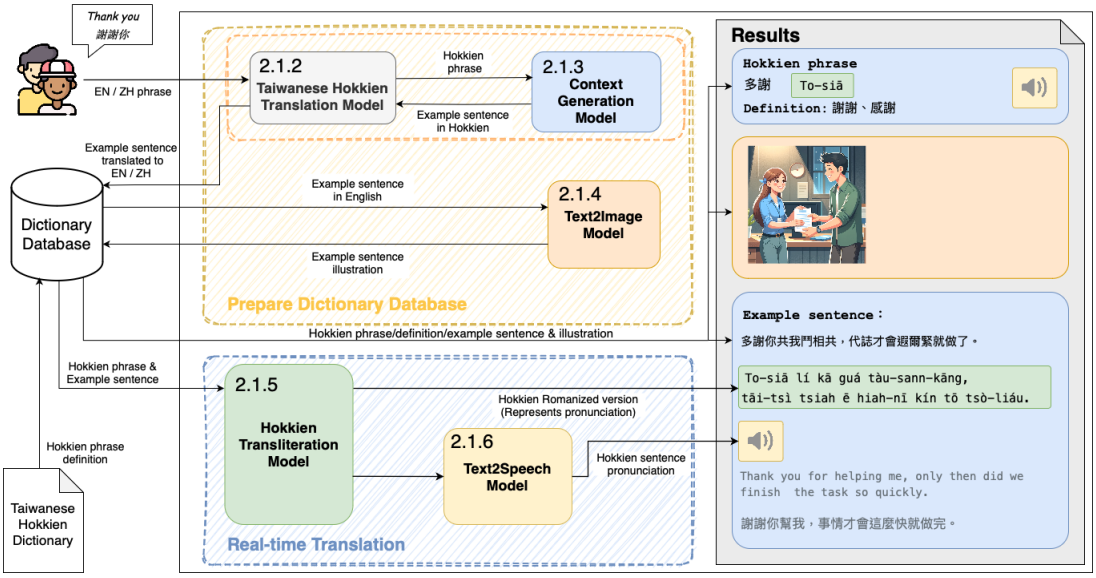

Figure 1: System Structure of ATAIGI. After translating phrases with the *Taiwanese Hokkien Translation Model* (Sec. 2.1.2, gray rounded box), we use the *Context Generator Model* (Sec. 2.1.3, blue rounded box), which is fine-tuned with Taiwanese Hokkien data, to obtain example sentences that contain the vocabulary using prompts. These example sentences, after translated back to English, are then inputted into *Text2Image Model* (Sec. 2.1.4, orange rounded box) which is used to generate images. Finally, the *Hokkien Transliteration Model* (Sec. 2.1.5, green rounded box) converts the words and sentences into Romanized pronunciation for *Text2Speech Model* (Sec. 2.1.6, yellow rounded box) to generate speech.

posed framework: a novel *Context Generation Model* that converts phrases into contextual sentences and a state-of-the-art *Hokkien Transliteration Model*. These custom-trained models addressed limitations in existing pre-trained tools for generating contextual content and accurate transliteration of Taiwanese Hokkien. The framework incorporates three additional models fine-tuned specifically for Taiwanese Hokkien: a *Taiwanese Hokkien Translation Model* (Lu et al., 2024), a *Text2Image Model*, and a *Text2Speech Model*, each carefully adapted to improve linguistic accuracy and performance within the Taiwanese Hokkien context. The above five models were human-evaluated and benchmarked to ensure their robustness and effectiveness in learning Taiwanese Hokkien.

We have developed a user-friendly learning application through these AI-powered generative models to bridge generational communication gaps for Taiwanese Hokkien. A comprehensive user study validated ATAIGI's efficacy, demonstrating significant improvements in Taiwanese Hokkien proficiency across diverse participants. High engagement and satisfaction ratings affirmed ATAIGI's potential for endangered language preservation. ATAIGI will carry the mission of cultural preservation and serve as a bridge for communications between generations. A video demonstration is available at https://youtu.be/fc409Q7jL2c.

2 ATaigi Language Application System

To use ATAIGI (Figure 1), the user first inputs an English or Chinese phrase, which is entered into the *Taiwanese Hokkien Dictionary* to retrieve its definition. Then the phrase is translated into Taiwanese Hokkien by the *Taiwanese Hokkien Translation Model* (Sec. 2.1.2). Meanwhile, the output of the *Translation Model* and *Context Generation Model* (Sec. 2.1.3) is Romanized by the *Hokkien Transliteration Model* (Sec. 2.1.5) and then converted into audible speech by the *Text2Speech Model* (Sec. 2.1.6) to assist users in correctly pronouncing and articulating phrases and sentences.

2.1 Backend AI Models

2.1.1 Datasets

There were five open-source datasets used to train the *Context Generation Model* and to build a dictionary for *Hokkien Transliteration Model* (Table 1). The *Context Generation Model* used *the Dictionary from Ministry of Education* for training and testing, and the *Hokkien Transliteration Model* used all datasets except *iCorpus* for building a dictio-

nary and used *iCorpus* for evaluation. Refer to Appendix B for more details.

Dataset	Entries Count	Training Model
Dictionary from Ministry of Education (for Context Generation Model)	16,863	Context Generator
Dictionary from Ministry of Education (for Hokkien Transliteration Model)	21,077	Hokkien Transliteration Model
Lyrics	322,449	
TAT	39,260	
Chhoetaigi	49,958	
iCorpus	450	

Table 1: Datasets for model training and testing. The datasets have been carefully curated to include a diverse range of topics (including official information, lyrics, to civics) to ensure that robust models are built. *Italicized* font is used to refer to dataset names.

2.1.2 Taiwanese Hokkien Translation Model

ATAIGI is integrated the *Taiwanese Hokkien Translation Model* introduced by Lu et al. (2024) is integrated to translate Mandarin and English from and to Taiwanese Hokkien. The model uses a collected Hokkien monolingual corpus and parallel datasets to train a *LLaMA2-based model* (Touvron et al., 2023). They conducted experiments to enhance dual translation performance between Taiwanese Hokkien with both English and Traditional Mandarin Chinese, resulting in the state-of-the-art translation[5] between Hokkien and these two high-resource languages in the open-source domain.

2.1.3 Context Generation Model

The *Context Generation Model* generates an example sentence in Taiwanese Hokkien containing the specified phrase for the user to understand the usage of the phrase in a sentence and contextual setting. The model utilizes supervised fine-tuning with a *pre-trained Hokkien 13B model* (Lu et al., 2024) based on *Llama2* model that is trained using over 630,000 tokens to generate sentences for given Taiwanese Hokkien phrases. A LoRA adapter (Hu et al., 2021) is trained on top of the pre-trained model using the dataset *Ministry of Education dictionary*[6], to better fine-tune the context generation model to output context that aligns more closely with daily usage of certain words. Refer to Appendix C for more details with the LoRA training.

We designed two types of prompts to generate contexts containing phrases. The first provides the phrases to the model and allows the model to generate example sentences composed of the phrases. However, due to a repetition error where one single word or phrases are repeated, we designed a second prompt to force the model to overcome this error by starting the phrase and preventing the model from taking over the context. Refer to Appendix D for more details.

2.1.4 Text2Image Model

To help users understand the context and meanings of the example sentence, we used a *Text2Image Model* that generates images that match the corresponding example sentence.

The *Text2Image Model* is based on the text-to-image diffusion model *DALL-E 3* from the OpenAI API. Since *DALL-E 3* only takes English prompts as input, the Hokkien contextual sentences generated by the *Context Generation Model* are first translated into English before being sent into the *Text2Image Model*. Furthermore, the API for the *DALL-E 3* model has safeguards that ensure the quality of the images generated, and no harmful content will be generated from the model.

To ensure that the *Text2Image Model* outputs appealing visual content for language learning, we requested for "comic" style in the prompts for the *DALL-E 3* model.

2.1.5 Hokkien Transliteration Model

The *Hokkien Transliteration Model* converts Hokkien Hanzi sentences into its complementary Hokkien Romanized phonetic representations so that the user can read the Hokkien pronunciation. This model functions by receiving input text in Hokkien Hanzi characters and producing the corresponding Romanized phonetic representation.

To achieve this, the implementation utilizes the jieba[7] segmentation tool to segment the input text into individual Hanzi words. These segmented words are then Romanized using a comprehensive dictionary consisting of 88,000 pairs of Hokkien Hanzi words and their corresponding Romanized phonetic representations. The dictionary is constructed from a dataset containing 432,744 pairs of sentences and encompasses both individual words and complete sentences. Adhering to the Taiwan Language Phonetic Alphabet, each Hanzi word's

[5] Achieved BLEU scores of 20.99 and 32.98 for translations from English and Traditional Mandarin Chinese to Taiwanese Hokkien Hanzi, respectively.

[6] The Ministry of Education dictionary

[7] jieba text segmentation: A popular Chinese word segmentation module tool.

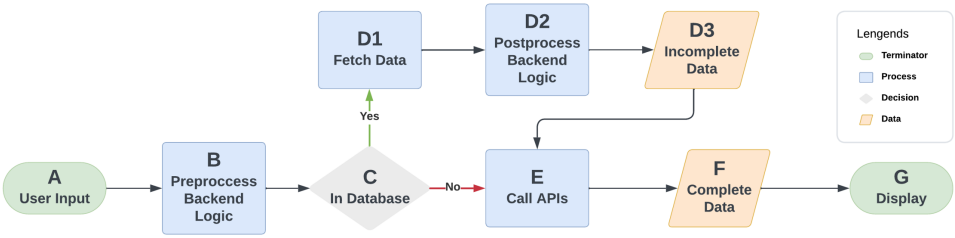

Figure 2: User Flow Diagram of ATAIGI showing Seamless User Interface and Effective Functionality.

representation is hyphenated for multi-syllable words, which is used to construct the dictionary. The selection process for the phonetic representation of each Hokkien Hanzi word prioritizes the most frequently occurring representation. Additionally, the system records each Hanzi character alongside its corresponding phonetic representation during the pairing process.

When a segmented word lacks a direct translation in the dictionary, a maximum matching algorithm is employed to translate the word. The phonetic representations are then hyphenated since they are within the same word.

2.1.6 Text2Speech Model

The *Text2Speech Model* is an extension of the Taiwanese across Taiwan (TAT) corpus (Liao et al., 2022), the first large-scale publicly available corpus of modern Taiwanese from across Taiwan. The authors used it as a dataset for model training and produced a high-quality Taiwanese language tool. By giving Hokkien phrases or sentences such as those generated by the *Hokkien Transliteration Model* and *Context Generation Model*, the *Text2Speech Model* generates audible Taiwanese Hokkien speech.

2.2 Frontend Framework Implementation

The deployed tool revolves around a REACT NATIVE web application, selected for its cross-platform compatibility with an emphasis on mobile users. Developed with EXPO in REACT NATIVE, NODE, NATIVE BASE (UI component library), and No-SQL database FIREBASE. The application is GITHUB auto-deployed with NETLIFY.

The user flow (Figure 2) with our application is as follows: upon inputting a phrase (with optional phrase suggestion) in English or Chinese (A), the back-end logic checks if the input exists in the database (B, C) and retrieves the corresponding translation, definitions, and example sentences (D1, D2, D3), along with API calls to the models described earlier (E) to retrieve visual representation and pronunciation (F). If the input does not exist, the application calls APIs (E) to retrieve the translation, visual representation, and pronunciation (F). The system logs unrecognized inputs for future evaluation and prevents the generation of misleading outputs for undefined words. At the bottom of the results page, users can provide positive or negative feedback using thumbs up or thumbs down buttons and leave comments on any app features to enhance user experience. More details can be found in Appendix E.1

The utilization of a database in our application serves a critical role in reducing the amount of time required for the user to get a response. Our database aims to prioritize efficiency and scalability to ensure possible future amendments to our model and our learning tool, bringing us to the database schema that contains two primary collections, `sentence` and `translation`, with details shown in Table 2.

3 Evaluation

Comprehensive evaluations were conducted for each model to ensure effective contribution and acceptable model performance. The evaluation is crucial for the correctness of each of the models by verifying that it: 1) functioned as intended, 2) generated the intended data, and 3) contributed effectively to the overarching objective.

Context Generation Model To ensure the *Context Generation Model* generates appropriate sentences, we evaluated its fluency, semantic correctness and sentence length. We utilize the state-of-the-art language model *GPT-4* to assess these aspects (Fu et al., 2023). Customized prompts included the phrase, its corresponding sentence in Hokkien and its definition (Appendix D). We used *GPT-4* to rate the fluency and semantic consistency between the generated sentence and the definition of the Hokkien phrase on a 10-point scale.

Collection	Field	Description
Sentence	sentences	Stores sentences in Traditional and Hokkien Chinese, and English.
	translationList	A list of unique document IDs in the Translations collection.
	imageURL	The URL of a pre-generated image (Figure 1) from the input phrase.
Translation	chineseInput	The input phrase in simplified or traditional Chinese.
	definitions	Hokkien explanations of the phrase.
	englishDefinitions	English explanations of the phrase
	englishInput	The input phrase in English.
	hokkienTranslation	The translated Hokkien phrase.
	sentence	A unique document ID in the Sentence collection.

Table 2: Database Schema for the ATAIGI with collection type and each name of the fields with descriptions.

Due to *GPT-4*'s limited performance with low-resource languages, we initially employ our *Taiwanese Hokkien Translation Model* to translate the generated sentences and definitions from Hokkien to English. Following this translation, we proceed with the evaluation.

The scores rated by *GPT-4* were then aggregated and averaged, resulting in a fluency score of 8.19 and a meaning score of 7.71. These scores reflect that *Context Generation Model* has strong performance, indicating that our *Context Generation Model* can produce sentences that are both linguistically fluent and semantically coherent. This evaluation method, using pre-eminent language models such as *GPT-4*, provides a standardized, objective and cost-effective approach to measuring the capabilities of our *Context Generation Model*. More importantly, the high fluency and meaning scores generated demonstrate the effectiveness of the *Context Generation Model*, which highlights its robustness in our framework.

Text2Image Model In order to evaluate the images, we make use of the HEIM evaluation method (Lee et al., 2023), where we selected two metrics from the HEIM evaluation method , aesthetics and alignment, supplemented by a human evaluation, to holistically assess the quality of the image generated in our system (Table 3).

For aesthetics evaluation, we employed the LAION Aesthetics metric (Schuhmann et al., 2022), to quantify visual appeal. Compared to a Stable Diffusion XL Model (SDXL 1.0), our chosen model has obtained an aesthetics score higher than the baseline SDXL model. This makes it a better choice compared to SDXL to output more aesthetically appealing images that complement our system.

To evaluate the alignment between the generated images and their corresponding sentence, we utilize the CLIP Score (Hessel et al., 2021) , indicating the degree of alignment between the visual content and the textual descriptions. Our chosen model has obtained a better CLIP Score than a standard SDXL model, making it a better choice than SDXL to output images that align better with the given context.

In addition to these automated metrics, we conducted a human evaluation on the *Text2Image Model* alongside the user study for our system. This human assessment provides a holistic judgement that complements the machine-derived scores mentioned previously. In the user study, our chosen model received a high rating of 4.6/5 on average, across all image component questions, demonstrating its effectiveness. Details on the user study will be elaborated in the **Qualitative Human Evaluation** section.

This study employs a multifaceted evaluation approach to comprehensively analyze the model's image generation performance. Based on the scores shown in Table 3, the DALL-E 3 model outperforms SDXL, and is selected for image generation in this work due to its strong performance.

Model	LAION (↑)	CLIP Score (↑)	Human Evaluation (↑)
SDXL 1.0	6.849	30.335	3.7/5
DALL-E 3 (ours)	**6.961**	**31.948**	**4.6 / 5**

Table 3: *Text2Image Model* assessed on HEIM metrics. **Bold** indicates the best performance in a metric. ↑ indicates the higher the better.

Hokkien Transliteration Model We selected 450 sentences from the *iCorpus* dataset to evaluate whether the *Hokkien Transliteration Model* can accurately translate the input sentences to its Romanized phonetic representation. We employed three evaluation metrics, BLEU score[8] (Papineni et al., 2002), chrF++ (Popović, 2015), and Word Error Rate (WER). Two distinct evaluation strategies were used. The first strategy involved adding

[8]BLEU is computed using $1-4$ gram precision, uniform weights, brevity penalty, and no smoothing by default.

spaces before and after hyphens to prevent multi-syllable words from being inaccurately considered as a single word in the BLEU score metric. The second strategy involved removing hyphens to examine the performance of our model on individual words. Based on these metrics and evaluation methods (Table 4), our approach demonstrates superior performance compared to both *GPT-4* and the *Hokkien LLaMA Translation Model* (Lu et al., 2024).

Model	With hyphen			Without hyphen		
	BLEU (↑)	chrF++ (↑)	WER (↓)	BLEU (↑)	chrF++ (↑)	WER (↓)
Ours	66.49	80.19	15.33	81.14	89.94	8.39
GPT-4	28.66	49.91	39.57	28.11	52.96	41.85
Hokkien LLaMA	44.42	61.74	31.16	57.01	70.81	26.73

Table 4: *Hokkien Transliteration Model* assessed on multiple metrics. **With hyphen** means adding spaces before and after the hyphen, while **Without hyphen** means removing the hyphen. **Bold** indicates the best performance in a metric. ↑ indicates the higher the better. ↓ indicates the lower the better.

Qualitative Human Evaluation The efficacy of ATAIGI was evaluated through a comprehensive study[9] involving 25 participants with diverse ages and proficiency levels. The assessment encompassed multiple dimensions, including usage patterns, functionality, learning outcomes, and user satisfaction (Table 5). Results indicated high user engagement (mean 10 sessions per week), with all subjects demonstrating enhanced language proficiency (mean increase of 18 points in test scores). Participants were also tested on their listening and reading comprehension skills using reference materials from official sources[10], further validating the learning outcomes. The overall satisfaction and unanimous intent to continue using ATAIGI highlight its effectiveness as a learning platform, emphasizing its potential for endangered language education and cultural heritage preservation across generations and proficiency levels.

4 Conclusion

Our ATAIGI is designed to teach Taiwanese Hokkien through phrases, example sentences, pictures, audio speech, and translations. The system is composed of five AI models, the most notable of which include an innovative *Context Generation Model* and high-quality *Hokkien Transliteration Model*. This system aims to prevent the extinction of low-resource languages and bridge the gap between different languages and generations. Empirical evaluation with heterogeneous participants substantiated ATAIGI's efficacy in enhancing Hokkien proficiency, confirming its potential for endangered language preservation. In the future, speakers of low-resource languages could use this system to overcome cross-language barriers and preserve their cultural heritage.

[9] ATaigi Survey Link
[10] Hokkien Test Link

Metric	Fluent	Beginner
Sample size	10	15
Age range	14-50+ years	
Weekly usage frequency	8.7 sessions	12.3 sessions
Core functionality efficacy	3.5 / 5	4.3 / 5
Perceived vocabulary acquisition	3.3 / 5	4.0 / 5
Cultural comprehension enhancement	3.0 / 5	3.6 / 5
Initial test score Average	65.4 / 100	44.8 / 100
Midterm test score range (First week)	75.2 / 100	60.5 / 100
Final test score range (One month)	82.1 / 100	73.3 / 100
Willing to use ATAIGI instead of Dictionary or ChatGPT	4.5 / 5	4.8 / 5
Perceived contribution to language preservation	4.7 / 5	4.9 / 5
Overall user experience and usability	4.75 / 5	
Overall satisfaction	4.5 / 5	

Table 5: User Study Evaluation Metrics of ATAIGI. Key performance indicators and user feedback are summarized from a questionnaire survey, with ratings on a 5-point scale. Results are divided into fluent users and beginners.

5 Limitations

The system architecture of ATAIGI is designed with high levels of modularity and flexibility, thus enabling seamless integration of state-of-the-art models as these become available. While the current image generation model may not represent the latest advancements, it provides a solid foundation for continuous improvement. Similarly, the Context Generation Model is evaluated using *GPT-4*, with Hokkien sentences translated into English due to *GPT-4*'s limited Hokkien proficiency. Despite the acknowledged tendency of *GPT-4* to produce high scores, it is important to note that the Hokkien transliteration model was initially evaluated using 450 sentences, thereby establishing a benchmark for future expansion.

The implementation of the tool framework may lead to higher storage costs, particularly in image storage. Assuming a worst-case scenario where each word in the Taiwanese Hokkien Language Common Words Dictionary requires a unique image, this would be approximately 16,000 images or 60 GB, which is a very large cost in terms of pre-computed storage. Currently, we are parsing manually generated CSV data to upload to FIREBASE to reduce load time and provide pre-computed results, which would not be feasible for larger datasets. In addition, FIREBASE, being a No-SQL database,

is suitable and scalable for the current use case. However, if the tool requires more complicated relational database features, it could lead to inconveniences. Transitioning to SQL databases paired with ORM such as PRISMA would provide better structure and long-term scalability and make the usage of No-SQL redundant. However, this transition would involve an initial increase in production time.

Ethics Statement

Ensuring the appropriateness and reliability of content is significant, especially within the context of an educational tool, where even minor deviations can yield substantial impacts. Consequently, this study has deployed rigorous human vetting protocols before incorporating example sentences into the database. The objective of this manual review process is to filter and eliminate any sentences or visuals that may depict violent or vulgar content. This screening procedure becomes particularly vital in situations where the model demonstrates a tendency to generate violent or inappropriate text and images. This action is pivotal in maintaining the ethical integrity of the application and ensuring the provision of suitable and constructive content.

The paper has been approved by Ontario Tech REB: 17995 and UofT REB: 00046686.

References

Zixin Chen. 2023. The study of hokkien with a comparison of the current hokkien situation in taiwan and the philippines. *Studies in Social Science Humanities*, 2:17–21.

Kosuke Doi, Katsuhito Sudoh, and Satoshi Nakamura. 2021. Large-scale English-Japanese simultaneous interpretation corpus: Construction and analyses with sentence-aligned data. In *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)*, pages 226–235, Bangkok, Thailand (online). Association for Computational Linguistics.

Jinlan Fu, See-Kiong Ng, Zhengbao Jiang, and Pengfei Liu. 2023. Gptscore: Evaluate as you desire.

Sourojit Ghosh and Aylin Caliskan. 2023. Chatgpt perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across bengali and five other low-resource languages. AIES '23, New York, NY, USA. Association for Computing Machinery.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. 2021. CLIPScore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7514–7528, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models.

Sunyoung Kim, Joobo Shim, and Jaechang Shim. 2023. A study on the utilization of openai chatgpt as a second language learning tool. *Journal of Multimedia Information System*, 10(1):79–88.

Tony Lee, Michihiro Yasunaga, Chenlin Meng, Yifan Mai, Joon Sung Park, Agrim Gupta, Yunzhi Zhang, Deepak Narayanan, Hannah Benita Teufel, Marco Bellagente, Minguk Kang, Taesung Park, Jure Leskovec, Jun-Yan Zhu, Li Fei-Fei, Jiajun Wu, Stefano Ermon, and Percy Liang. 2023. Holistic evaluation of text-to-image models.

Yuan-Fu Liao, Chia-Yu Chang, Hak-Khiam Tiun, Huang-Lan Su, Hui-Lu Khoo, Jane S. Tsay, Le-Kun Tan, Peter Kang, Tsun-guan Thiann, Un-Gian Iunn, Jyh-Her Yang, and Chih-Neng Liang. 2020. Formosa speech recognition challenge 2020 and taiwanese across taiwan corpus. In *2020 23rd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, pages 65–70.

Yuan-Fu Liao, Jane S. Tsay, Peter Kang, Hui-Lu Khoo, Le-Kun Tan, Li-Chen Chang, Un-Gian Iunn, Huang-Lan Su, Tsun-Guan Thiann, Hak-Khiam Tiun, and Su-Lian Liao. 2022. Taiwanese across taiwan corpus and its applications. In *2022 25th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, pages 1–5.

Bo-Han Lu, Yi-Hsuan Lin, En-Shiun Annie Lee, and Richard Tzong-Han Tsai. 2024. Enhancing hokkien dual translation by exploring and standardizing of four writing systems.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Roxana Rebolledo Font de la Vall and Fabian Gonzalez Araya. 2023. Exploring the benefits and challenges of ai-language learning tools. *International Journal of Social Sciences and Humanities Invention*, 10:7569–7576.

Cécile Robin, Gautham Vadakkekara Suresh, Víctor Rodriguez-Doncel, John P. McCrae, and Paul Buitelaar. 2022. Linghub2: Language resource discovery tool for language technologies. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6352–6360, Marseille, France. European Language Resources Association.

Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models.

Lee Schwartz, Takako Aikawa, and Michel Pahud. 2004. Dynamic language learning tools. In *InSTIL/ICALL2004 Symposium on Computer-Assisted Language Learning*.

David Stap and Ali Araabi. 2023. ChatGPT is not a good indigenous translator. In *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 163–167, Toronto, Canada. Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Jin Ha Woo and Heeyoul Choi. 2021. Systematic review for ai-based language learning tools.

Appendix

A Background and Preliminaries

Taiwanese Hokkien, also known as Hokkien, Hoklo, Taigi, Southern Min, or Min Nan, is a branch of the Southern Min dialects. It has evolved independently of those in Fujian due to historical and geographical separation and has been influenced by indigenous languages, Dutch, and Japanese (Liao et al., 2020). The writing system for Taiwanese Hokkien is currently categorized into four types: Traditional Chinese with additional characters (Hanzi), two types of Romanization systems known as Pe̍h-ōe-jī (POJ) and kàu-iok-pōo Tâi-lô (KIP), as well as mixed representation. This paper exclusively uses KIP for Romanization.

B Dataset Details

The Dictionary from Ministry of Education serves as a key resource for training our *Context Generation Model* and building a dictionary for *Hokkien Transliteration Model*. This authoritative dictionary provides crucial word information, phonetic readings, definitions, and example sentences. From its $16,863$ usable entries, a random split for training and test data yielded $9,446$ entries for training and $7,417$ for testing, focusing on words with example sentences and excluding idiomatic expressions. *Hokkien Transliteration Model* uses $21,077$ entries for its dictionary.

The *lyrics* dataset was compiled from Hokkien songs shared in Facebook communities[11], totaling $322,449$ entries.

The *TAT* dataset consists of $39,260$ entries sourced from speech transcription texts from the Taiwanese Across Taiwan (Liao et al., 2020) speech recognition competitions.

The *ChhoeTaigi*[12] dataset integrates multiple datasets, compiled and digitized for easier retrieval of Hokkien words. In total, we used $49,958$ entries.

iCorpus[13] is created by Academia Sinica. dataset is composed of news articles from diverse domains. We selected 450 entries from this corpus to evaluate our *Hokkien Transliteration Model*.

C Training Details for Context Generation Model

We implemented parameter-efficient fine-tuning via LoRA with the specifications shown below:

Learning Rate	1e-4
LoRA Rank	64
Dropout Rate	0.05

Table 6: Training settings for Context Generation Model

[11] The compilation of Taiwanese Hokkien songs by private citizens, featuring Hokkien lyrics and its Romanized forms.

[12] https://github.com/ChhoeTaigi/ChhoeTaigiDatabase

[13] The Taiwanese Hokkien News Corpus

D Prompts Used in Context Generation Model

Our ATAIGI uses two types of prompts to generate data for different scenarios. During the model training process, we input Taiwanese Hokkien phrases at `{phrase}`. Prompt '`"{BOS}[INST]\n{phrase}\n[/INST]"`' generates sentences for each word, most of which are fluent and reasonable. However, some words are not included in the generated sentences or do not form reasonable sentences after several attempts. Therefore, prompt '`"{BOS}{phrase}"`' is used to process this part of the text. It is important to note that the generated sentence will always start with the specified word due to a prompt limitation. Therefore, this characteristic is considered of lower priority.

On the other hand, we use *GPT3.5-turbo* via prompt to evaluate the output of the *Context Generation Model*. The evaluation template is generated by *GPT4*. The input should include the Taiwanese Hokkien phrase in `word`, the Taiwanese Hokkien sentence in `sentence`, and the Chinese definition of the phrase in `definition`. The evaluate prompt template is as follows.

```
Word: {word}
Sentence: {sentence}
Word definition: {definition}
On a scale of 1-10, how fluent and
logical does this sentence appear to be
? In the context of this sentence?, how
match does the meaning of this word and
its defined explanation? On a scale of
1-10
```

E Frontend Details

E.1 Application Structure Overview

The application's architecture is modular, focusing on specific user interface components and interaction handling. The main components are **User Interface** that is designed for easy navigation and interaction, it displays translation results along with pronunciation and images. **API Calls** facilitates communications with the connection endpoints to the models. **Back-end Logic** efficiently manipulates data to store and display it from the database.

User Interface The user interface (Figure 3 in Appendix) contains five pages: landing, home, loading, result, and settings.

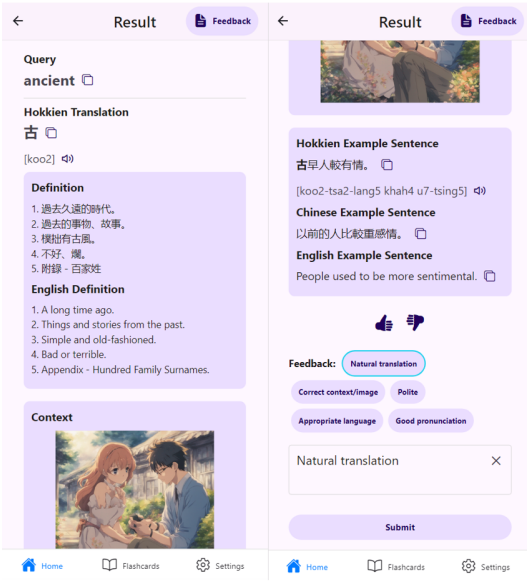

Figure 3: User Interface. Develop user-friendly interfaces to enhance usability and user experience.

API Calls The REACT NATIVE application utilizes AXIOS[14], a promise-based HTTP client, to manage API calls. This choice allows for an efficient and straightforward way to send asynchronous HTTP requests to REST (Representational State Transfer) endpoints and perform CRUD (CREATE, READ, UPDATE, and DELETE) AXIOS simplifies the process of fetching data and handling responses, making the app's data interaction smooth and reliable.

Back-end Logic The back-end logic mainly handles the following functionalities, store data from EXCEL files to FIREBASE, CRUD for database collections as needed and enables the user flow as discussed previously (Figure 2).

E.2 User Interface

Frontend user interface of the application (Figure 3)

[14] Axios Project Page